

OM of: US-09-303-518D-653 to: SwissProt\_40:\* out\_format : pfs  
Date: Jun 30, 2002 9:10 AM

About: Results were produced by the Gencore software, version 4.5  
Copyright (c) 1993-2000 Compugen Ltd.

```

Command line parameters:
-o=/cgn2.1/USPRO.spool/US09303518/runat_28062002_142714.4362/app_query.fasta.1.23501
-model=framat-n2p.model -DEV=xlh
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blusun62
-TRANS=human40.cdi -LIST=100 -DOCALLIG=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFM=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09303518.@cgn1.1.440 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPY -WAIT -THREADS=1

```

Search information block:  
Query: US-09-303-518D-653

```
Database sequences: 105224
Database length: 38719550
Search time (sec): 217.960000
```

score\_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
SwissProt_40:IGAL_HAEIN			1614.00	1416.21	2,2e-72	! P42782 haemophilus influenzae
SwissProt_40:IGGA_HAEIN			1587.50	1392.04	4.5e-71	! P44695 haemophilus influenzae
SwissProt_40:IGGA_HAEIN			1587.00	1392.08	4.5e-71	! P44585 haemophilus influenzae
SwissProt_40:IGGA_HAEIN			1583.50	1386.47	7.1e-71	! P45384 haemophilus influenzae
SwissProt_40:IGGA_HAEIN			1565.50	1373.49	5.4e-70	! P09790 nisseria gonorrhoeae
SwissProt_40:IGGA_HAEIN			1516.50	1328.68	1.4e-67	! P45386 haemophilus influenzae
SwissProt_40:YPAJ_ECOLI			359.00	309.32	9.9e-11	! P52143 escherichia coli
SwissProt_40:YPAJ_ECOLI			292.50	252.60	1.8e-07	! P54508 escherichia coli
SwissProt_40:HLVA_PROMI			277.50	237.40	1.0e-06	! P16466 proteus mirabilis
SwissProt_40:PERT_BOBPA			275.50	240.19	1.0e-06	! P24328 bordetella parapertussis
SwissProt_40:PERT_BOBPA			274.00	238.96	1.4e-06	! P03035 bordetella bronchiseptica
SwissProt_40:PERT_BOBPA			270.00	235.45	2.2e-06	! P146283 bordetella pertussis
SwissProt_40:OMPA_RICRY			267.50	228.23	3.1e-06	! P69689 r outer membrane protein
SwissProt_40:OMPA_RICRY			261.50	220.29	6.3e-06	! P15921 rickettsia rickettsii
SwissProt_40:OMPA_RICRY			250.50	211.49	2.2e-05	! P05627 rickettsia conorii
SwissProt_40:ATGA_ECOLI			247.50	212.68	2.9e-05	! P03155 escherichia coli
SwissProt_40:ATGA_ECOLI			245.00	206.93	4.0e-05	! P1953 salmonella typhimurium
SwissProt_40:VEJO_ECOLI			243.00	212.08	4.7e-05	! P33924 escherichia coli
SwissProt_40:YDAB_ECOLI			238.50	200.99	8.4e-05	! P33666 escherichia coli
SwissProt_40:YDAB_ECOLI			237.50	205.66	8.9e-05	! P33180 escherichia coli
SwissProt_40:OMPB_RICRY			237.00	201.74	9.3e-05	! P06653 r outer membrane protein
SwissProt_40:OMPB_RICRY			237.00	201.28	9.8e-05	! P053020 r outer membrane protein
SwissProt_40:YDCK_ECOLI			235.00	201.40	0.0001	! P096251 escherichia coli
SwissProt_40:PM21_CHLPI			224.00	190.05	0.0004	! P02605 chlamydia pneumoniae
SwissProt_40:OMPB_RICRY			223.50	189.38	0.0005	! P053047 r outer membrane protein
SwissProt_40:PHAB_BOBPA			216.50	176.64	0.0011	! P12255 bordetella pertussis
SwissProt_40:HXAZ_HAEIN			215.50	187.22	0.0011	! P45354 haemophilus influenzae
SwissProt_40:HLVA_SEMA			213.00	180.36	0.0015	! P15320 serratia marcescens
SwissProt_40:PMBP_CHLPI			209.00	178.79	0.0023	! P09689 chlamydia pneumoniae
SwissProt_40:V100_YEAST			206.50	179.00	0.0029	! P02692 saccharomyces cerevisiae
SwissProt_40:YACA_HELPI			203.50	173.85	0.0042	! P55981 helicobacter pylori
SwissProt_40:PMOP_RICRY			203.50	173.78	0.0042	! P14914 rickettsia rickettsii
SwissProt_40:PM20_CHLPI			203.50	171.65	0.0043	! P09712 chlamydia pneumoniae
SwissProt_40:WAPB_BACSU			202.50	167.94	0.0050	! P09812 chlamydia pneumoniae
SwissProt_40:VAC3_HELPI			200.50	171.07	0.0050	! P048253 bacillus subtilis
SwissProt_40:PMGP_CHLPI			199.00	171.93	0.0069	! P084879 chlamydia trachomatis
SwissProt_40:VCG3_BPPI			199.00	166.55	0.0071	! P07067 bacteriophage t2
SwissProt_40:AMTV_YEAST			199.00	169.39	0.0071	! P08640 saccharomyces cerevisiae

[illegible]

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK368 / SEROTYPE B:
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Theegersen H.C., Kilian M.;
RT "Cloning and sequencing of the Immunoglobulin A1 protease gene (iga)
  of Haemophilus influenzae serotype b.";
RL Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MUTAGENESIS OF SER-288.
RC STRAIN-HK368 / SEROTYPE B:
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
  influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-1-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; X64357; CAA45708.1; -.
DR EMBL; M87492; AAA24969.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPRO00710; Iga_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINIS; PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1008 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1009 1541 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT MUTAGEN 288 288 S->T: LOSS OF ACTIVITY.
SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

```

## alignment\_scores:

Quality: 1614.00 Length: 1716  
Ratio: 1.732 Gaps: 48  
Percent Similarity: 54.312 Percent Identity: 26.865

## alignment\_block:

US-09-303-518d-653 x IGA1\_HAEIN ..

Align seg 1/1 to: IGA1\_HAEIN from: 1 to: 1541

```

64 CGCTTCGCGCCGCTTACTTACCATATGCGTTCGCGCATTCCTGCC 113
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
5  LysPheLysLeuAsnPhelLeuAlaLeuThrValAlaLeuThrPr 21
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
114 CCAAGCCGGGGGCGACACACTTATTGGCATCACTACCAATGCTATC 163
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
21  oTyThrGuaAlaAlaLeuValArgAspValAspValArgInLlePheA 38
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
164 GCGAGTTTCCGAAATTAAGGCAAGTTTGCGTGGGGCGAAGATATT 213
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
38  rGAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

```

```

214 GAGGTTTACAAACAAAAGGAGTGTGTCGCCAAATCGATGACGAAGC 263
   ||| ::::::::::: ::::| ::::| ::::|
55  LeuValLysAspLysAsnLysAspLeuGlyThrAlaLeuProAsnG1 71
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
264 C...CCGATGATGATTTTCTGTGTATCG...CGTAAAGCGGTGGCG 307
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
71  yLleProMetLleAspPheSerValAlaAspValAspLysArgLleAla 88
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
308 CATTGGCGGGCGATCATATATATGTAGCGTGCACAT...AACGGCGC 354
   ::::| ::::| ::::| ::::| ::::| ::::|
88  hrLeuLleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
355 TATAACATGTTGATTTGGT...GCGAGGAGGAACAA 389
   ::::::::::: ::::| ::::| ::::| ::::|
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGly 421
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
390 TCCCGATCAGCAGCCG...TTTCTTACCAATG 421
   ::::| ::::| ::::| ::::|
121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
422 TGAACAAATATATTAAAGCAGGACTAACGCG... 456
   ::::| ::::| ::::| ::::|
138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThrThr 154
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
457 .....CATCTTATGGCGGATTCATATCATGCGGCTTGCA 494
   ::::| ::::| ::::| ::::|
155 GluAspLeuThrGlnLysArgValGluAspTyrTyrMetProArgLeuAs 171
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
495 CAAATTTGTCAGAGATGACAGACCTGTGAG...ATGACCAGTTATATG 541
   ::::| ::::| ::::| ::::|
171 pLysPheValThrCysValAlaProLleGluAlaSerThrAlaSerSer 188
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
542 ATGGGTGGAATACCGTCGATTTAAATTAATACCTGATCGTGTGAATC 591
   ::::| ::::| ::::| ::::|
188 sPrlaGlyThrTyrAsnAspLysAsnLysTyrProAlaPheValArgLeu 204
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
592 GGAAGCAGCAGACAAATATGGCGTGTGATGAGAGAA... 630
   ::::| ::::| ::::| ::::|
205 GlySerCylSerGlnPheLleTyrLysLysGlyAspAsnTyrSerLeuI 221
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
631 .CCCAATACCGCGGAAGTTCAATATCAT...ATTGCAACG 667
   ::::| ::::| ::::| ::::|
221 eLeuAsnAsnHisGluValGlyLysAsnLeuLysLeuValGlyAspA 238
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
668 CATATGTTGGTTCGCGGCGCAATACCTTGCAACAAATGATGACAGG 717
   ::::| ::::| ::::| ::::|
238 lAtyThrTyrGlyLleAlaGlyThrProTyrLysValAsnHisGluAsn 254
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
718 GGTGGCAGCTCAACTAGTACGCAAAAATTAACATGAC...CCATA 764
   ::::| ::::| ::::| ::::|
255 AsnGlyLeuLleGlyPheGlyAsnSerLysGluGlnHisSerAspProLy 271
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
765 TGGTTTTTTACCAACAGGA...GGCTCATTTGGCGACA 799
   ::::| ::::| ::::| ::::|
271 sGlyLleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
800 GTGGCTCAGCAATGTTATCTATGATGCCCCAAGCAAGAGTGGTAAAT 849
   ::::| ::::| ::::| ::::|
288 erGlySerProLeuPheValTyrAspArgGlyLysGlyLysTrpLeuPhe 304
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
850 AATGGGTTATTCGAACAGCAACCCCTATATAGAAAACCAATGGCTT 899
   ::::| ::::| ::::| ::::|
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer... 318
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
900 CCAAGTACTGTCGTAAGATGG...TTCTATGATGAATCTTTGCTGGAG 946
   ::::| ::::| ::::| ::::|
319 .....TrpGlnLysTrpAsnLleTyrLysSerGlnPheThrLysA 332
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
947 ATACCCATTCAGATATCTACGACACATCAACAAATGGAATATCTTTT 996
   ::::| ::::| ::::| ::::|
332 sPvalLeu... 334

```

```

997 AACGACATATATATGCGCAGCAAAAATCGATGCCAAATTAACACTA 1046
    |||.....|||.....:
335 .....AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349
1047 TTCTCTACCTTTAGATTAAAAACAGCAACGCTT.....CAAT 1084
    |||.....:
349 rSerTrpSerSerAsnGlyLysThrSerThrIleThrGlyGlyLys 366
1085 TGTTAATGTTCTTTATCCGACAGCAGCAAGAACCTGTTATCATGCT 1134
    |||.....:
366 eLeuAsnValAlaSplLeuAlaSplGlyLysAspLysPro..... 378
1135 GCAGGTGGGGTCAACAGTTATCGACCCAGCTGAATTAATGAGAAATAT 1184
    |||.....:
379 .....AsnHisGlyLysSerVa 384
1185 TTCTTTATGACAAAGAAAGTGAATGTATNCTTACCAGCAACATCA 1234
    |||.....:
384 ThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnHisIleA 399
1235 ACCAAGCGCGCGCGCTTTGTATTTTGAAGGTAATTTACGGTC...TCG 1281
    |||.....:
399 sPcInGlyAlaGlyLysLeuPhePheGlyGlyAspTyGlyValLysGly 415
1282 CCTAAACACAGCAACAGTGGCAGCGCGCGCTTCATATCATGTGATGG 1331
    |||.....:
416 ThrSerAspAsnThrTrpLysGlyAlaGlyValSerValAlaGly 432
1332 CAGTACCGCTTCTGGAAGATGAACGGCGCTGCAACAGCCGCTGTCCA 1381
    |||.....:
432 LysThrValThrTrpLysValHisAsnProGlnTyAspArgLeuAla 449
1382 AAATGCGCAAAAGCAGCGTGTGTTCAAGCAAAAGGAAACCAAGGC 1431
    |||.....:
449 ySleGlyLysGlyThrLeuIleValGlyGlyThrGlyAspAsnLysGly 465
1432 TCGGTACGCGTGGCGCAGCGTAACTATCTTATGATCAGCAGCGCGACA 1481
    |||.....:
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnG 482
1482 TCAAGGCAAAAACAGCCTTAGTGAATCGCTTGTGTCAGCGCGCAAGG 1531
    |||.....:
482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyAsn 498
1532 GGACGCGCAATGATGATCGATATATCAATCAACCCGACAACTCAT 1581
    |||.....:
498 eThrLeuValLeuAsnAspAspLysGlnValAlaAspProAsnSerIleTy 514
1582 TTGGCGCTTTCGCGCGCAGCGTGGATTGAACGGGCAATTCGCTTCGT 1631
    |||.....:
515 PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531
1632 CCACGCGCATTCGAATACCGATGAGAGGGCGATGTTGTCAACCAACATC 1681
    |||.....:
531 eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAlaAsnHisAsnM 548
1682 AAGACAGAAGATCCACCGTTACCATTAAGCAATTAAGATATTAATCT... 1728
    |||.....:
548 eThrAsnAlaSerAsnIleThrIleThrGlyGlySerLeuIleThrAsp 564
1728 ..... 1728
565 ProAsnThrIleThrProTyAsnIleAspAlaProAspGluAspAsnPr 581
1728 ..... 1728
581 oTyAlaPheArgArgIleLysAspGlyGlyLeuLeuTyLeuAsnLeuG 598
1728 ..... 1728
598 LuAsnTyThrTyThrAlaLeuArgGlyAlaAlaSerThrArgSerGlu 614
1729 .....ACAACCGCAATTAAC..... 1743

```

---

```

615 LeuProLysAsnSerGlyGlySerAsnGluAsnTrpLeuTyMetGly 631
    |||.....:
1744 .....AACACTTGGATGAGCAAAA 1762
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluA 648
1763 AAGAAATTTGGCTACAAACGTTGGTTGGCGAGAAAGCAATGCAACCAACG 1812
    |||.....:
648 rMetAsnGlyPheAsnGlyTyThrPheGlyGlyGlyGlyGlyGlyGlyGly 663
1813 AACGGCGGCTCAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1862
    |||.....:
664 AsnGlyLysLeuAsnValThrPheLysGlyLysSerGlyGlyGlyGlyGly 680
1863 ACTGCTTCCGGGAGCAATTAATTAACGGCATATTCAGCCCAACCAACG 1912
    |||.....:
680 eLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGlyLysG 697
1913 GCAAACTGTTTTCAGCGCAGACCGACCGACCGACCGACCGACCGACCGAC 1962
    |||.....:
697 LysThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
1963 GGAAGCGGTTGTCAAAATGGAAGT.....ATCCACAGAGAGAAAT 2006
    |||.....:
714 GlyLysSerSerThrLysAspProHisPheAlaGlyAsnAsnGlyVal 730
2007 CGTGGGAGCAGACGATGTGATCGACCGCAGCATTTAAACGGGAAACCTCC 2056
    |||.....:
730 lValValGluAspAspTrpLysAsnArgAsnPheLysAlaThrThMetA 747
2057 ATATTGAGGCGGACAGACCGGTGTTTC...CGCATGTTGCCAAAGTC 2103
    |||.....:
747 sValThrGlyAsnAlaSerLeuTySerGlyArgAsnValAlaAsnIle 763
2104 GAAGCGCATTTGGCATTAACCAATCACGCCCAAGCACTTTCGGTGTCCG 2153
    |||.....:
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTy 779
2154 ACCGCATCAAAAGCCACACATGTGATCGCTGCGGAGCGGCTGTGA 2203
    |||.....:
780 ....LysThrGlyAspThrValCysValArgSerAspTyThrGlyTy 795
2204 CAAGTGTACCGGAAACCAATTAACCGCATGAAGTGAATGCTTCATTTG 2253
    |||.....:
795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810
2254 AGCAGACCGCAATCAGAGCAATGTCAAGCTTGGCGATCAGCTCATTTT 2303
    |||.....:
811 AsnProThrAsnLeuArgGlyAsnValAlaSerLeuThrGlySerAlaAsnPh 827
2304 AAATCTCACAGGACTTGCACACTCAACAGCGCATTCATTAGTGCAGGGGAG 2353
    |||.....:
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843
2354 ACACGCACTATACGTTACCGCAGCAACGCCACCAAAACGGCAACTCAGC 2403
    |||.....:
843 sNserGlnValArgLeuThrGluAsn..... 851
2404 CTCGTGGCAATGGCCACAGCAACATTTATCAAGCCACATTAAGCGCAA 2453
    |||.....:
851 ..... 851
2454 CACATCGGCTTGGACAAATGCTTATTATTAAGCAACACAGCGGTAC 2503
    |||.....:
2504 AAAAGCGAGCTCTGACGTTTCCGACACGCTAAGCAAAACGTAAGCAT 2553
    |||.....:
851 ..... 851
2554 TCGGACTCAAGGCAATGTCTCCCTAGCCGATTAAGCAGTATTCGATTT 2603

```

```
851 ..... 851
2604 TGAACAGCGCGCTTACCGGAAATTCAGCGCGGCAAGATACGGCAT 2653
851 ..... 851
2654 TACACTTAAAGACAGCAATGAGCGTCGCGGCAAGAAATAGCC 2703
852 ..... SerHisThrIleuThrGlyAsnSerAspValHis 863
2704 AATTAAACCTTGACACGCCACATTCATCATTCGCGCTATCGACA 2753
864 GlnLeuAspLeuAlaAsnGlnHisIleHisLeuAsnSerAlaAspAsnSe 880
2754 CGATGCGGACGCGGCAACCGGACAGTCGCGAGATGCGCGCGCGCC 2803
880 rAsn..... 881
2804 GTTCGCGCGCTTCCCTATTATCGGTTACGCGCGCAACTTCGCGAGAAATCC 2853
882 ..... AsnValThr..... 884
2854 CGTTTCAACGCTGACGCTAAACGCAATTCGAACGCTCAGGACATC 2903
885 LysTyrAsnThrIleuThrValAsn...SerLeuSerGlyAsnGlySerPh 900
2904 CCGCTTTATGTCGAACTCTCGGCTACCGGCAAGCGCAATTCGAAGCTGG 2953
900 eTyTyrIleuThrAspLeuSerAsnLysGlnLysAspLysValValT 917
2954 CGGAAGTTCCGAGAGGCACTTACCTTGGCTGCAACATACCGGCAAC 3003
917 hTyrSerAlaThrGlyAsnPhetThrLeuGlnValAlaAspLysThrGly 933
3004 GAACCGGTAGTCTCGAACAATTGACGTAGTGAAAGGAAACACACAC 3053
934 GluPro...AsnHisAsnGlnLeuThrLeuPheAspAla.....SerLys 947
3054 ACCGCTGTCCGAAATCTTAATTCACCTCGCAAAACGACACGTGCATG 3103
947 salaglnatrgasphHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGGCGCATGGCGTTATCAGCTTATCCGCAAGACGCGAGTTCCGCGCTG 3153
964 euGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980
3154 CATAATCCG..... 3162
981 TyrAsnProGlnValGlnLysArgAsnGlnThrValAspThrThrAsnIl 997
3162 ..... 3162
997 eTrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 ..... 3162
1014 snGlnGlnIleAlaArgValAspGlnAlaProValProProProAlaPro 1030
3162 ..... 3162
1031 AlaThrProSerGlnThrThrGlnThrValAlaGlnAsnSerLysGlnG 1047
3162 ..... 3162
1047 userLysThrValGlnLysAsnGlnAlaAspAlaThrGlnThrThrAlaG 1064
3163 ..... GTCAAAGACACAGAG 3177
1064 lAsnArgGlnValAlaLysGlnAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTTCGACAAACTCGGCAAGCGGGA...GAACAGAGCGCGCTTGAC 3224
1081 GlnThrAsnGlnValAlaGlnSerGlySerGlnThrLysGlnThrGlnTh 1097
3225 GGCAAAACAGGCACAACTTGCCCGCAAAACACAAACAGCGGCAAAACAGCAACG 3274
1097 rThrGlnThrLysGlnThrAlaThrValGlnLysGlnLysAlaLysV 1114
3275 CGCAAGC.....CTTGACGCGCTGATTGCG 3300
1114 alGlnThrGlnLysThrGlnGlnValProLysValThrSerGlnValSer 1130
3301 GCCGGCGCAATGCCACCGCAAAAGCGAAAGTTCGCCAACCAGCGCCG 3350
1131 ProLysGlnGlnLysSerGlnThrValGlnProGlnAlaGlnProAlaAr 1147
3351 CGACGCAAGCGGGAATGCGCGCATTCATTCAGCGGCAAGAGAGAAA 3400
1147 g.....GlnAsnAspProThrValAsnIleLysGlnProGlns 1160
3401 AACGGGTGCAGCGCGATTAAGACACCGCCTTG...GCCAAACAGCGGAA 3447
1160 eGlnThrAsnThrThrAlaAspThrGlnGlnProAlaLysGlnThrSer 1176
3448 GCG..... 3450
1177 SerAsnValGlnGlnProValThrGlnSerThrThrValAsnThrGlyAs 1193
3451 .....GAACCGCGCGGCTTACCGCGCTTCC 3478
1193 nSerValValGlnAsnProGlnAsnThrThrProAlaThrThrGln...P 1209
3479 CC.....CGCGCGCGCGCGC 3492
1209 rThrValAsnSerGlnSerSerAsnLysProLysAsnAlaGlnHisArgArg 1225
3493 GCCCGCGGGAATTGGCG...CAACCGCAGCGCCCAACCGCAACCGCAAC 3539
1226 SerValArgSerValProHisAsnValGlnProAlaThrThrSerSerAs 1242
3540 GCAGCGC.....GACCTGATCAGCGCTTATGCCAATA 3571
1242 nAspArgSerThrValAlaLeuLysAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTGAGTGAAATTTCCGCC.....ACGCTCACAGC 3606
1259 lValIleuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGCGGTACGAGCAATTCGACCGCGTG...TTGCCGAGACGCG 3653
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGlnLysAsnGlnGln 1292
3654 CCGCAACGCGGTTTGACAAACGCGCATCCGGACACCAACACATACCGTT 3703
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309
3704 CGCAAGATTTCCGCGCTTACCGCCACAAACCGACCTGCGCCAAATCGCT 3753
1309 eSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAGAAAACCTCGGACAGCGG...CGGTCGCGCATCTCGTTTGGCA 3800
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyAlaValPheThrTy 1342
3801 CAACCGACCGGAAACACTTCGACGACGCGCATCCGACACTCGGACGCGC 3850
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrLeuA 1359
3851 TTGCCACAGCGTTCGCGCAATACGCGCATCGGACGTTGCGACATC 3900
1359 lAsnValAsnPhetYrSerLysTyrGlnAlaAspAsnHisThrPyrLys 1375
3901 GGCATCAGCGCGGCGCGGCTTTAGTAGAGCGGACCTTCACAGCGCAT 3950
1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnH 1392
```



```

3951 CAGAGCGAAATCCGCCGCGTCTGCTATTCAGCATTCAGCAAGAT 4000
1392 sasnAlaIysPheAlaIargHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409
4001 ACCGCGCAGTTTCGGCGGATTCGATTCAGCAACCGCATTCGGCGCAGC 4050
1409 ysAlaPheAsnLeuGlyAsnPhelGlyLeuThrProLeuValGlyValarg 1425
4051 CGCTATTTCGTCGCAAAAGCGATTCAGTACGAAAGCAAGTAAATTCGC 4100
1426 TySerTyrLeuSerAsnAlaPheAlaLeuAspGlnAlaArgIleTyr 1442
4101 CACCCCGCGCTTCGATTCACCGCTACCGCGCGGCGCATTAAGCGAGAT 4150
1442 sValAsnProIleSerValIleThrAlaPheAlaGlnValAsnLeuSerT 1459
4151 ATTCAATTCACACCGCGCAGACATTCATTCACCGCTTATTGAGCGTG 4200
1459 yThrTyrHisLeuGlyGlu...PheSerValThrProLeuLeuSerAla 1474
4201 TCCTATACGATCGCGCTCGCGCAAGTCCGACAGCGCGGCTCAATACCGC 4250
1475 ArgTyr...AspAlaAsnGlnIleSerGlyLysIleAsnValAsnGlyTyr 1490
4251 CGTATTGGCGCAGATTCGCGCAAAACCGCAGTGGCGGATGGCGTAA 4300
1490 rAspPheAlaTyrAsnValGlnAsnGlnGlnIleThrAlaGlyLeuL 1507
4301 AGCCGCAATTCAGAGTTTCACGCTGTCCCTACCGCTCGCGCGCGCAAG 4350
1507 ySleuLysTyrHisAsnValLysLysSerLeuIleGlyGlyLeuThrLys 1523
4351 GGGCGCGCATTCGAGCGCAGCAGACAGCGCGGCGCATCAATTCAGGTAC 4398
1524 AlaIysGlnAlaGlnLysLysIleThrAlaGlnLeuLysLeuSerPhe 1539

seq_name: SwissProt_40:IGA0_HAEIN

seq_documentation_block:
ID IGA0_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE D;
RA Wright A., Fishman Y., Tai F., Plant A.G.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ueteyback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.

```

```

CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X59800; NOT_ANNOTATED_CDS.
CC EMBL: U32779; AAC22651.1; -.
CC DR MEROPS; S06.001; -.
CC DR TIGR; H10990; -.
CC DR InterPro; IPR000710; IGA_S6.
CC DR Pfam; PF02395; IGA1; 1.
CC DR PRINTS; PR00921; IGASERPTASE.
CC DR KMW Complete; Serine protease; Transmembrane; Zymogen; Signal;
CC KW Complete proteome.
CC FT SIGNAL 1 25
CC FT CHAIN 26 1014
CC FT PROPEP 1015 1694
CC FT ACT_SITE 288 288
CC FT CONFLICT 253 254
CC FT CONFLICT 272 272
CC FT CONFLICT 464 464
CC FT CONFLICT 866 866
CC FT CONFLICT 1036 1036
CC FT CONFLICT 1074 1074
CC FT CONFLICT 1421 1421
CC FT CONFLICT 1545 1545
CC FT CONFLICT 1545 1545
CC FT SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

alignment_scores:
Quality: 1587 50 Length: 1859
Ratio: 1.683 Gaps: 48
Percent Similarity: 50.726 Percent Identity: 25.282

alignment_block:
US-09-303-518D-653 x IGA0_HAEIN ..
Align seg 1/1 to: IGA0_HAEIN from: 1 to: 1694
64 CGCTTCGCGCGCTTACTAGCATTCGCTGCTCGCATTCGCGC 113
114 CCAAGCGCGCGCGCATTCATTCGCGCATTCACATTCATTCATTC 163
21 CTYrThGlnAlaLeuValArgAspAspValAspIlePhe 38
164 GCGCATTCGCGCAAAATTAAGCAAGTTCGCTGCGCGCAAGATTC 213
38 rgsAspPheAlaGlnLysGlyArgPheSerValAlaIleThrAsnVal 54
214 GAGCTTACACAAAAAAGGAGGAGTGTGCGCAAAATTCATGACGA 263
55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
264 C...CGATGATGATTTTCGTGTGTCG...CGTACGCGCGTGGCG 307
71 yIleProMetIleAspPheSerValAlaValAspAlaLysArgIleAla 88

```

```

308 CATGGCGGCGATCATATATATTTGAGCGGACAT...ACGCGCGC 354
      ::::: ||||| ::::: |||||
88 hrlEUILeasNProGIrValValGIyValLysHisValSerAsnGIy 104
      ::::: ||||| ::::: |||||
355 TATACAAATGTTGATTGGT.....GCGGAGGACGAA 389
      ::::: ||||| ::::: |||||
105 ValSerGIUleuHisPheGIyAsnLeuAsnGIyAsnMetAsnAsnGIyAs 121
      ::::: ||||| ::::: |||||
390 TCCCGATGAGACCGC.....TTTCTTACCAAAATG 421
      ::::: ||||| ::::: |||||
121 nAlaLysSerHisArgAspValSerSerGIUasArgTyrPheSerV 138
      ::::: ||||| ::::: |||||
422 TGAAGAAGAAATATATTAACGAGGACTAACGC..... 456
      ::::: ||||| ::::: |||||
138 aGIUlyAsnGIUlyrProThrLysLeuAsnGIyLysAlaValThrThr 154
      ::::: ||||| ::::: |||||
457 .....CATCCTTATGGCGGCGATTCATATGCGCGCTTGCA 494
      ::::: ||||| ::::: |||||
155 GluAspGIUThrGIUlyAsArgArgGIUAspTyrTyrMetProArgLeuAs 171
      ::::: ||||| ::::: |||||
495 CAAATTTGTACAGATGCGAAGCACTGTGAG...ATGACAGTTATATG 541
      ::::: ||||| ::::: |||||
171 pLysPheValThrGIUValAlaProIleGIUAlaSerThrAlaSerSerA 188
      ::::: ||||| ::::: |||||
542 ATGGGTGCAATACCGCTGATTAAATAATACCTGATCGTTCGATC 591
      ::::: ||||| ::::: |||||
188 sPAlaGIyThrTyrAsnAspGIUAsnLysTyrProAlaPheValArgLeu 204
      ::::: ||||| ::::: |||||
592 GGAGCAGCAGACAAATATGGCGGTGATGAAGCA..... 630
      ::::: ||||| ::::: |||||
205 GlysSerGIySerGIUProIleTyrLysLysGIyAspAsnTyrSerLeuI 221
      ::::: ||||| ::::: |||||
631 .CCCAATACCGCGAAGTTCATATCAT.....ATTGCAACG 667
      ::::: ||||| ::::: |||||
221 eLysAsnAsnHisGIUValGIyGIyAsnAsnLeuLysLeuValGIyAspA 238
      ::::: ||||| ::::: |||||
668 CATATTTCTGGCTCGTGGCGCATACCTTTGCACAAATGATGACAGT 717
      ::::: ||||| ::::: |||||
238 lAtyThrTyrGIyLleAlaGIyThrProTyrLysValAlaHisGIUAsn 254
      ::::: ||||| ::::: |||||
718 GGTGCGCAGACTAGTAGTAGCAAAATAATTAACATATAC...CCATA 764
      ::::: ||||| ::::: |||||
255 AsnGIyLeuIleGIyPheGIyAsnSerLysGIUAsnHisSerAspProLy 271
      ::::: ||||| ::::: |||||
765 TGGTTTTTACCACACAGA.....GGCTCATTTGGCGCA 799
      ::::: ||||| ::::: |||||
271 sGIyLysLeuSerGIUAspProLeuThrAsnTyrAlaValLeuGIyAspS 288
      ::::: ||||| ::::: |||||
800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTAT 849
      ::::: ||||| ::::: |||||
288 eGIySerProLeuPheValTyrAspArgGIUlyGIyLysTyrLeuPhe 304
      ::::: ||||| ::::: |||||
850 AATGGGATATTGCAACAGCGCAACCCCTATATAGAAAAGCAATGGCGT 899
      ::::: ||||| ::::: |||||
305 LeuGIySerTyrAspPheThrAlaGIyTyrAsnLysLysSerTyrPGL 321
      ::::: ||||| ::::: |||||
900 CCAGTACTGCTAAAGATTGGTCTATGATCAAAATCTTGTGCGAGATA 949
      ::::: ||||| ::::: |||||
321 uTrpAsnLysTyrLysProGIUleuAlaLysThrValLeuAspLysAsp 338
      ::::: ||||| ::::: |||||
950 CCCATTACGATTTACGACACACATCAAAATGGGAATACTTTTAAAC 999
      ::::: ||||| ::::: |||||
338 hT..... 338
1000 GACAAATAATATGGCGAGAAAATCGATGCCAACAATATAACTATTC 1049
      ::::: ||||| ::::: |||||
339 .....AlaGIySerLeuThrGIySerAsnThrGIyAs 350
      ::::: ||||| ::::: |||||
1050 TCTACCTTATAGATTAAAAACAGACCGTT.....CAATTGT 1087
      ::::: ||||| ::::: |||||
350 nTrpAsnProThrGIyLysThrSerValIleSerAsnGIySerGIUserL 367
      ::::: ||||| ::::: |||||
1088 TTAATGTTTCTTTATTCGAGACAGCAAGAACCTGTTATCATGCTGCA 1137

```

```

      ::::: ||||| ::::: |||||
367 eLysnValAspLeuPheAspSerSerGIUAsp..... 377
      ::::: ||||| ::::: |||||
1138 GTGGGGGTACACAGTTATGACCCAGACATGAATATGAGAAATATTC 1187
      ::::: ||||| ::::: |||||
378 .....ThrAspSerLysLysAsnHisGIyLysSerValTh 390
      ::::: ||||| ::::: |||||
1188 CTTTATTTGACAAAGAAAGTGAATTGATTTACTTACACACATCAAC 1237
      ::::: ||||| ::::: |||||
390 rLeu.....ArgGIySerGIyThrLeuThrLeuAsnAsnIleAspG 405
      ::::: ||||| ::::: |||||
1238 AAGCGCGGGCGGCTTTGATTTTGGAGGTAATTTACGGTC...TCGCT 1284
      ::::: ||||| ::::: |||||
405 lndGIyAlaGIyGIyLeuPhePheGIUGIyAspTyrGIUValLysGIyThr 421
      ::::: ||||| ::::: |||||
1285 AAAAACACAGCAAGCGTGAAGCGGCGGCTTCATATGATGATGACG 1334
      ::::: ||||| ::::: |||||
422 SerAspSerThrThrTyrPlySGlyAlaGIyValSerValAlaAspGIy 438
      ::::: ||||| ::::: |||||
1335 TACCGTTACTTGAAGAAAGTGAAGCGCGTGCACAAAGCGCTGCCAAA 1384
      ::::: ||||| ::::: |||||
438 sThrValThrTyrPlyValHisAsnProLysSerAspArgLeuAlaLysI 455
      ::::: ||||| ::::: |||||
1385 TCGGCAAAAGGCGAGCTCGTGTTCACAGCAAGGGAAGAAACCAAGCTCG 1434
      ::::: ||||| ::::: |||||
455 lGIyLysGIyThrLeuIleValGIUGIySGlyLysGIUAsnLysGIySer 471
      ::::: ||||| ::::: |||||
1435 GTGACGTGGCGCGACGTAAGTCAATCTTAGATCAGCAGCGGACGATCA 1484
      ::::: ||||| ::::: |||||
472 LeuLysValGIyAspGIyThrValIleLeuLysGIUAlaAspAlaAs 488
      ::::: ||||| ::::: |||||
1485 AGGCAAAAACCAAGCCTTAGTGAATCGGCTTGGTCAGCGGACGGGGA 1534
      ::::: ||||| ::::: |||||
488 nAsnLysValLysAlaPheSerGIUValGIyLysGIyValArgSerT 505
      ::::: ||||| ::::: |||||
1535 CGGTGCACTGAATGCCGATATCACTGATCAACCCGCAACACTCTATTTC 1584
      ::::: ||||| ::::: |||||
505 hTrValIleuAsnAspAspLysGIUValAspProAsnSerIleTyrPhe 521
      ::::: ||||| ::::: |||||
1585 GGGTTTCGCGCGGCGAGTTTGGATTGAACGGGCGATTCGCTTGTTC 1634
      ::::: ||||| ::::: |||||
522 GIyPheArgGIyGIyArgLeuAspAlaAsnGIyAsnAsnLeuThrPheGI 538
      ::::: ||||| ::::: |||||
1635 CCGCATTCAAATACCGATGAGAGGGCGATGATGTCACACCAATCAAG 1684
      ::::: ||||| ::::: |||||
538 uHisIleArgAsnIleAspAspGIyAlaArgLeuValAsnHisAsnThrS 555
      ::::: ||||| ::::: |||||
1685 ACAAGAAATCCACCGTTACCATTCAGCGCAATTAACATATTACT..... 1728
      ::::: ||||| ::::: |||||
555 eLysThrSerThrValThrIleThrGIyGIUserLeuIleThrAspPro 571
      ::::: ||||| ::::: |||||
1728 ..... 1728
572 AsnThrIleThrProTyrAsnIleAspAlaProAspGIUAspAsnProTy 588
      ::::: ||||| ::::: |||||
1728 ..... 1728
588 rAlaPheArgArgIleLysAspGIyGIyGIULeuTyrLeuAsnLeuGIUa 605
      ::::: ||||| ::::: |||||
1728 ..... 1728
605 snTyrThrTyrTyrAlaLeuArgLysGIyAlaSerThrArgSerGIUleu 621
      ::::: ||||| ::::: |||||
1729 .....ACAACCGGCAATAAC..... 1743
      ::::: ||||| ::::: |||||
622 ProLysAsnSerGIyGIUserAsnGIUAsnTyrPleuTyrMetGIyLysTh 638
      ::::: ||||| ::::: |||||
1744 .....AACAATTGGATGACAAAAAG 1765
      ::::: ||||| ::::: |||||
638 rSerAspGIUAlaLysArgAsnValMetAsnHisIleAsnAsnGIUArgM 655
      ::::: ||||| ::::: |||||
1766 AAATTCCTCAACGCGTGGTGGGAGAGAAAGATGCAACCAAAAGCAAC 1815
      ::::: ||||| ::::: |||||

```

```
655 etAsnGlyPheAsnGlyTyrPheGlyGluGluGly...LysAsnAsn 670
1816 GGGCGGCTGATCTGTAATTACCAACCGGAGAGAGCGGATCGCACTTTACT 1865
|||||
671 GlyAsnLeuAsnValThrPheLysGlySerGluGlnAsnArgPheLe 687
1866 GCTTTCGGCGGAGCAAAATTTAAAGCGCAATATCAACGCAACAGCGCA 1915
|||||
687 uLeuThrGlyGlyThrAsnLeuAsnGlyAsnLeuLysValGluLysGlyT 704
1916 AACTGTTTTCAGCGGAGACCGGACAGCGGACCGCTACATCATTTAGGA 1965
|||||
704 hLeuPheLeuSerGlyArgProThrProHisAlaArgSerIleAlaGly 720
1966 AGCGGGTGTCAAAATGGAAGT....ATCCCAACAAGGAATGCT 2009
|||
721 IleSerSerThrLysLysAspGlnHisPheAlaGluAsnAsnGluValVa 737
2010 GTGGGACACAGATTTGATCGACCGGACATTTAAACGGGAAAACTTCATA 2059
|
737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnV 754
2060 TTCAGGCGCGACAGCGGTTGTC...CGCAATGTTCCAAAGTGGAA 2106
|||||
754 aThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThr 770
2107 GCGCATTTGGCATTTAAGCAATCAGCGCCAGAGATTTTCGGTGTGGCAC 2156
|||||
771 SerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr.... 785
2157 GCATCAAAAGCCACACATCTGTACAGCTTCGAGTGAAGCGGTCGACAA 2206
|||
786 LysAlaGlyAspThrValLysValArgSerAspTyrThrGlyTyrValT 802
2207 GTTGACCGAAAAAACCATTTACCGGACGATTAAGTATGCTTCATTTAGC 2256
|||||
802 hCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsn 817
2257 AAGACCGACATCAGAGCAATGTACGCTTCGCGATTCACGCTCATTTAAA 2306
|||||
818 AlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
2307 TCTCACAGAGACTTGGCACACTCAAGCGCAATCTTACTGACAGCGGAGACA 2356
828 ..... 828
2357 CGCACTATACGGTTACGGCAACGCCACCAAAAGCGCAACCTGACCTC 2406
828 ..... 828
2407 GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAGCGCAAC 2456
|||||
829 ...GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrI 844
2457 ATCGGCTCGGACATGCTTCATTTAATCAAGCAACAGCGGTCACAAA 2506
|||||
844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn..... 857
2507 ACGGCACTGTACGCTTCCGCAACAGCTAAGCAAAAGTAAGCATTTCC 2556
857 ..... 857
2557 GCACTCACGCAATGTCTCCCTAGCCGATAGCGCATTTTCATTTTGA 2606
857 ..... 857
2607 AAACAGCGCTTTACCGGAAAAATCAGCGGCGGCAAGGATACGGCATTTAC 2656
857 ..... 857
2657 ACTTAAAGACAGCGAATGACGCTCGCGTCGGCAGCGAATTAGCAAT 2706
|||||
858 .....SerHisTrpHisLeuThrGlyAspSerAsnValAsnGln 870

2707 TTAACCTTGACACAGCCACCATTTACACTCAATTCGCCCTATGACACGA 2756
|||||
871 LeuAsnLeuAspLysGlnHisIleHisLeuAsnAla.....GlnAsn 885
2757 TGCGGAGGCGCGCAACCGGCACTGCGGCAAGATTCGCGCGCGCGCGCTT 2806
|||||
885 pAlaAsnLysValThrThr..... 891
2807 GCGCGCTTCCTCATTTATTCGTTACGCCGCCAACTTCGGCAGATCCCGT 2856
891 ..... 891
2857 TTCAACACGCTGACGGTAAACGGCAATTTGACAGGTCAGGAAATTCG 2906
|||||
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnIleSerPheT 907
2907 CTTTATGTCGAACCTCTTCGGCTACCGGACGCGCAATTTGACCTGGCGG 2956
|||||
907 rTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValAlaThrL 924
924 ySerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGlyGlu 940
2957 AAAGTTCGGAAGGCATTTACACCTTGCGTGTCAACAATACCGGCAAGAA 3006
|||||
3007 CCGCTAGTCTCGACCAATTGACCGGTGTGGAAGGAAGAACACACACACC 3056
|||||
941 ProThrLys...AsnGluLeuThrLeuPheAspAlaSerSerAsnAlaThr 955
3057 GCTGTCCGAAAATCTTAATTTCAACCCGTGCAAAAGCAACACGATGATCGC 3106
|||||
956 ...ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAspLeuG 971
3107 GCGATGCGCTTATCAGCTTATCCGCAAGACGCGGAGTTCCGCCCTGCAT 3156
|||||
971 lYalaTrpLysTyrLysLeuArgAsnValAsnGlyArgAspLeuTyr 987
3157 AATCCG..... 3162
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004
3162 ..... 3162
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021
3162 ..... 3162
1021 lGluIleAlaArgValGluThrProValProProAlaProAlaThr 1037
3162 ..... 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGluSerLy 1054
3163 .....GTCAAGACACAGAGCTTTCGCAACAACCTCGCAAGCGG 3202
|||||
1054 sThrValGluLysAsnGlnGlnAspAlaThrGluThrThrAlaGlnAsnG 1071
3203 GAGAA..... 3207
1071 lGluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208 .....ACAGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGl 1104
3213 GCGCGCCTTGACGCA..... 3228
1104 uIleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGluL 1121
3229 .....AAA 3231
1121 yAspGluIleGlnGluAlaProGlnMetAlaSerGluThrSerProLys 1137
```



01	SEQUENCE	1545 AA; 170627 MW; 3E9D753988F6D478 CRC64;
02	ACT_SITE	292
03	PROPEP	1013 1545
04	FT	292
05	FT	1013
06	FT	1545
07	CHAIN	1 25
08	SIGNAL	1 25
09	KW	Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
10	DR	PRINTS: PR00921; IGASERPTASE.
11	DR	InterPro: IPR000710; IGA_S6.
12	DR	Pfam: PF02395; IGAL1; 1.
13	DR	EMBL: M87490; AAA24967.1; .
14	CC	MEROPS: S06.001; .
15	CC	CC
16	CC	CC
17	CC	CC
18	CC	CC
19	CC	CC
20	CC	CC
21	CC	CC
22	CC	CC
23	CC	CC
24	CC	CC
25	CC	CC
26	CC	CC
27	CC	CC
28	CC	CC
29	CC	CC
30	CC	CC
31	CC	CC
32	CC	CC
33	CC	CC
34	CC	CC
35	CC	CC
36	CC	CC
37	CC	CC
38	CC	CC
39	CC	CC
40	CC	CC
41	CC	CC
42	CC	CC
43	CC	CC
44	CC	CC
45	CC	CC
46	CC	CC
47	CC	CC
48	CC	CC
49	CC	CC
50	CC	CC
51	CC	CC
52	CC	CC
53	CC	CC
54	CC	CC
55	CC	CC
56	CC	CC
57	CC	CC
58	CC	CC
59	CC	CC
60	CC	CC
61	CC	CC
62	CC	CC
63	CC	CC
64	CC	CC
65	CC	CC
66	CC	CC
67	CC	CC
68	CC	CC
69	CC	CC
70	CC	CC
71	CC	CC
72	CC	CC
73	CC	CC
74	CC	CC
75	CC	CC
76	CC	CC
77	CC	CC
78	CC	CC
79	CC	CC
80	CC	CC
81	CC	CC
82	CC	CC
83	CC	CC
84	CC	CC
85	CC	CC
86	CC	CC
87	CC	CC
88	CC	CC
89	CC	CC
90	CC	CC
91	CC	CC
92	CC	CC
93	CC	CC
94	CC	CC
95	CC	CC
96	CC	CC
97	CC	CC
98	CC	CC
99	CC	CC
100	CC	CC
101	CC	CC
102	CC	CC
103	CC	CC
104	CC	CC
105	CC	CC
106	CC	CC
107	CC	CC
108	CC	CC
109	CC	CC
110	CC	CC
111	CC	CC
112	CC	CC
113	CC	CC
114	CC	CC
115	CC	CC
116	CC	CC
117	CC	CC
118	CC	CC
119	CC	CC
120	CC	CC
121	CC	CC
122	CC	CC
123	CC	CC
124	CC	CC
125	CC	CC
126	CC	CC
127	CC	CC
128	CC	CC
129	CC	CC
130	CC	CC
131	CC	CC
132	CC	CC
133	CC	CC
134	CC	CC
135	CC	CC
136	CC	CC
137	CC	CC
138	CC	CC
139	CC	CC
140	CC	CC
141	CC	CC
142	CC	CC
143	CC	CC
144	CC	CC
145	CC	CC
146	CC	CC
147	CC	CC
148	CC	CC
149	CC	CC
150	CC	CC
151	CC	CC
152	CC	CC
153	CC	CC
154	CC	CC
155	CC	CC
156	CC	CC
157	CC	CC
158	CC	CC
159	CC	CC
160	CC	CC
161	CC	CC
162		

```
alignment_scores:
  Quality: 1587.00
  Ratio: 1.710
  Percent Similarity: 53.860
  Length: 1723
  Gaps: 47
  Percent Identity: 26.291
```

alignment\_block:

US-09-303-518D-653 x IGA3\_HAEIN .

Align seg 1/1 to: TGA3\_HAEIN from: 1 to: 1545

64 CCGTCTTCCGCGCTTACTTACCAATATGCCGTGCTGGCATTTGCG 113  
 5 LysPheLysLeuAsnPhelLeuLeuThrValAlaIryAlaLeuThrPr 21  
 114 CCAGCCCGGGGGGACACACTTATTTGGCATCACTACCAATATCATC 163  
 21 GYrrYrrGuaAlaAlaLeuValaIArgAspAspValAspIrrLllePheA 38  
 164 GCACATTTCCGCAATAATAAGCAAGTTTCGACGCGGGCGCAAGATATT 213  
 38 rGAspPheAlaGlnAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54  
 214 GAGGTTTAACACAAAAAGGGGAGGTGTGTCGCAATGATGACGAGAAGC 263  
 55 GluValaIArgAspLysAsnAsnIArgProLeuLylsAsnValLeuProAsnG1 71  
 264 C...CCGATGATGATTTTTTCTGTGGTATGCG...CGTAAAGCGCGTGGCG 307

[illegible]



```
2633 GCGGGCGCAAGATACGGCATTTACATTAAAGACGGAATGACGCTG 2682
856 ..... SerHisTriphHisLeu 860
2683 CCGTGGGCGACGGAATTAGCAATTAAACCTTGACAAGCCGACCATAC 2732
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlnHisLeu 877
2733 ACTCAATTCCGGCTATCGACAGATGGCGGCGGCAACCGGCATCG 2782
877 SLeuAsnSerAlaAspAsnSerAsn..... 885
2783 CGGCAGATGCGCGCGCGCGCTTCGCCCTTATTCCTATTCGCTACG 2832
886 ..... AsnValThr 888
2833 CCGCCAACTTCGGCAGAAATCCCGTTTACACGCTGACGGTAACGGCA 2882
889 ..... LysTyrAsnThrLeuThrValAsn...Se 897
2883 ATTGAACGCTCAGGAGACATTCGGCTTATGCGAATCTTCGGCTACG 2932
897 rLeuSerGlnAsnGlnSerPheTyrTyrLeuThrAspLeuSerAsnLys 914
2933 GCAGCGCAAAATTGAAGCTGGCGGAAGTCCGAAGCCTTACACCTTG 2982
914 InGlnAspValValValThrLysSerAlaThrGlnLysAsnPheThrLeu 930
2983 GGTGCACAAATACCGCAACGACCCCTAAGTCCGAGCAATGACGCT 3032
931 GlnValAlaAspLysThrGlnLysPro...AsnHisAsnGlnLeuThrLe 946
3033 AGTGGAGGAAAGACACACACCGCTGCCAAATCTTAATTACACCC 3082
946 uPheAspAla....SerLysAlaGlnArgAspHisLeuAsnValSerL 961
3083 TGCAAACGAAACGCTGATCGCGCGCATGGCTTATCAGCTTATCCGC 3132
961 euValGlnAsnThrValAspLeuGlnAlaThrLysTyrLysLeuArgAsn 977
3133 AAGACGCGGAGTCCGCTGCATTAATCCG..... 3162
978 ValAsnGlnArgTyrAspLeuTyrAsnProGlnValGlnLysArgAsnG 994
3162 ..... 3162
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011
3162 ..... 3162
1011 alProSerValProSerAsnAsnGlnGlnIleAlaArgValAspGlnAla 1027
3162 ..... 3162
1028 ProValProProProAlaProAlaThrProSerGlnThrThrGlnThrVa 1044
3162 ..... 3162
1044 lAlaGlnAsnSerLysGlnGlnSerLysThrValGlnLysAsnGlnGlnAla 1061
3162 ..... 3162
1061 spAlaThrGlnThrThrAlaGlnAsnArgLysValAlaLysGlnValAlaLys 1077
3163 .....GTCAAGAAAGAACAGAGCTTCGACAAATCCGCAAGGCGGA.. 3204
1078 SerAsnValLysAlaAsnThrGlnThrAsnGlnValAlaGlnSerGlySe 1094
3205 .GAAACAGAGGCGCGCTTGACGGCAAAACAGGACCACTTCCGCCCAAC 3253
1094 rGlnThrLysGlnThrGlnThrThrGlnThrLysGlnThrAlaThrValG 1111
3254 AACAGGCGGAAAAAGACACCGCGCAAGC..... 3282
```

---

```
1111 LulysGlnGlnLysAlaLysValGlnThrGlnLysThrGlnGlnValPro 1127
1127 ..... 1127
3283 ...CTTGACGCGCTGATTCGGCGCGCGCAATGCCAACCGCAAGGCGAGA 3329
1128 LysValThrSerGlnValSerProLysGlnGlnSerGlnThrValGln 1144
3330 AAGTGTGCCGAAACCGCGCGCGGACGAGCGGGGAAATGCCGCGATT 3379
1144 nProGlnAlaGlnProAlaArg.....GluAsnAspProThrV 1157
3380 TGCAGCGCGAGAGAGAAAAACGGGTGACGCGGATAAGACACCGCC 3429
1157 alAsnIleLysGlnProGlnSerGlnThrAsnThrThrAlaAspThrGln 1173
3430 TTG...GCGAAACAGCGCGAAGCG..... 3450
1174 GlnProAlaLysGlnThrSerSerAsnValGlnProValThrGlnSe 1190
3451 .....GAAACCC 3457
1190 rThrThrValAsnThrGlnAsnSerValValGluAsnProGluAsnThrT 1207
3458 GCGCGGCTACACCGCCTTCCCG..... 3480
1207 hrProAlaThrThrGln...ProThrValAsnSerGlnSerSerAsnLys 1222
3481 .....CGCGCGCGCGCGCGCGCGGATTTCCG...CAACGCA 3518
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValG 1239
3519 GCCCAACCGCAACCCCAACCGCAGCGC.....GACC 3550
1239 uProAlaThrThrSerSerAsnAspThrSerThrValAlaLeuCysAspL 1256
3551 TGATCAGCCGTTATGCCAATAGCGGTTTGAGTAATTTCCGCGC..... 3594
1256 euThrSerThrAsnThrAsnAlaValLeuSerSerAlaThrGlnAlaLysAla 1272
3595 .....ACGCTCAACAGCGCTTTCCGCGGTACAGAGCAATTTGAGCCG 3635
1273 GlnPheValAlaLeuAsnValGlnLysAlaValSerGlnHisIleSerG 1289
3636 CGTG...TTTGGCGAAGACCGCGCAACCGCGTTTGACACAGCGGATCC 3682
1289 nLeuGlnMetAsnAsnGlnLysGlnTyrAsnValThrValSerAsnThrS 1306
3683 GGGACACCAACACATACCGTTCCGAAGATTCGCGCGCTACCGCCAAAC 3732
1306 erMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLys 1322
3733 ACCGACCTGCGCAATCGGTATGCAGAAAAACCTCGGACGCGG...CG 3779
1323 SerThrGlnThrGlnLeuGlnTyrPaspGlnThrIleSerAsnAsnValG 1339
3780 CGTGGCATCCGCTTTTGCGACACCGGACCGGAACACCTTCGACGACG 3829
1339 nLeuGlnGlnValAlaPheThrTyrValAlaArgAsnSerAsnAsnPheAspLysA 1356
3830 GCATTCGCAACTCGGCGCGCTTGCCACAGGTGCTTTCGGGCAATAC 3879
1356 lThrSerLysAsnThrLeuAlaGlnValAsnPheThrSerLysTyrTyr 1372
3880 GGCATTCGCAAGTTGCACATCGGCATCAGCGCGCGCGGCTTTTAAGT 3929
1373 AlaAspAsnHisThrPyrLeuGlnLysLeuGlnTyrGlnLysPheGln 1389
3930 CGGACGCTTTCAGACGCGCATGACAGGCAAAATCCGCGCGCGCGTGC 3979
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406
3980 ATTACGCAATTCAGGCAAGATACCGCGCAGGTTTCGGGATTCGCGATC 4029
```



```

1406 InPhegIleuThraIaGlyLysAlaPheAsnLeuGIyAsnPhgIyIle 1422
4030 GAACCGCATGGCGGCAACGGCGTATTGTCGCAAAAAGGATACCG 4079
1423 ThrProIleValIaIArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439
4080 ATACGAAACGTCAATATCGCCACCGCGGCGCTTCGATTCACAGCTAC 4129
1439 AleuAspIlnAlaIArgIleLysValAsnProIleSerValIyThrAlaP 1456
4130 GCGCGGCGATTAAGCAGCATATTCATTCACAAACCGCGCCACACATTC 4179
1456 heaIaGlnValaIAspLeuSerTyrThrIyHisIeAsnValGlu...PheSer 1471
4180 ATCAGCGCTTATTTGCGCTGCTCATACGATCGCGCTTCGCGCAAGT 4229
1472 ValThrProIleLeuSerAlaIArgTyr...AspAlaAsnGlnGlySerGI 1487
4230 CCGAAGCGCGCTCAATACCGCGCGTATTCGCGCAGATTTCGCAAAACC 4279
1487 LysIleAsnValaIAsnGlyTyrAspPheAlaIyAsnValaIaIAsnGlnG 1504
4280 GAGATGGCGAATGGGCGTAAACCGCGCAATCAAAAGTTTCACGCTGTC 4329
1504 InGlnIyTrAsnAlaGlyLeuLysLeuLysTyrHisAsnValIyLysLeuSer 1520
4330 CTCACGCTGCGCGCGCGCGCGCGCATTCGCAAGCGCAGACAGCGC 4379
1521 LeuIleGlyLysLeuThraLysGlnAlaGlyLysGlnLysThrAl 1537
4380 GGGCATCAATTAGCTAC 4398
1537 agLLeuLysLeuSerPhe 1543

seq_name: SwissProt_40: IGA2_HAEIN

seq_documentation_block:
ID IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (ICAI protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBL_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / SEROTYPE B.
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases."
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87489; AAA24966.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; zymogen; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1014 POTENTIAL.
FT PROPEP 1015 1702 IMMUNOGLOBULIN A1 PROTEASE.
FT ACT_SITE 288 288 HELPER PEPTIDE (POTENTIAL).
FT DOMAIN 1109 1124 PROBABLE.
FT 2 x 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
FT REPEAT 1109 1116 K.
FT REPEAT 1117 1124 1.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

alignment_scores:
Quality: 1583.50 Length: 1867
Ratio: 1.679 Gaps: 48
Percent Similarity: 50.509 Percent Identity: 25.174

alignment_block:
US-09-303-518D-653 x IGA2_HAEIN ..
Align seg 1/1 to: IGA2_HAEIN from: 1 to: 1702

64 CGCTTCGCGCGCGCTTACTTAGCCATATGCTGTGCGCATTCGCC 113
5 LysPheLysLeuAsnPhelIeAlaLeuThrValaIaIaLeuThrPr 21
114 CCAAGCGCGGCGGCGACACACTATTTCGCGCATCCTCCATATAC 163
21 IyTrInGlnAlaIaLeuValaIArgAspAspValaIAspTyrGlnIlePheA 38
164 GCACATTGGCGGCAAAATAAGCAAGTTTGACGCGCGCGCAAGATAT 213
38 rGAspPheAlaGlnaIAsnLysGlyArgPheSerValaIaIaThrAsnVal 54
214 GAGGTTTACACAAAAAAGGGAGTGTGTCGCGCAATTCATGCGAAGC 263
55 GluValaIArgAspLysAsnAsnHisSerLeuGlyAsnValaIleuProAsnG 71
264 C...CCGATGATGATTTTCTGTGATGCG...CCTAACGGCGTGGCG 307
71 yIleProMetIleAspPheSerValaIAspValaIAspLysArgIleAlaI 88
308 CATTGGCGGCGGATCATATATTTGAGCGTGGACAT...AACGGCGC 354
88 hrLeuIleAsnProGlnTyrValaIaIaGlyValaIyHisValaISerAsnGly 104
355 TATAACAATGTGATTTTGTG...GCGGAGGAGAACAA 389
105 ValSerGlnLeuAsnIspheGlyAsnLeuAsnGlyAsnMetAlaAsnLys 121
390 TCCGCATCAGCAGCG...TTTCTTACCAATG 421
121 naAspLysSerHisArgAspValSerSerGlnGlnaIAsnArgTyrPheSerV 138
422 TGAAGAAGAAATAATTAAGACGAGGACTAACGCG... 456
138 alGluLysAsnGlnIyTyrProThrLysLeuAsnGlnLysAlaValThrThr 154
457 .....CATCCTATGCGCGCATATATCATATGCGCGTTTGA 494
155 GluAspGlnThrGlnLysArgArgGlnAspTyrTyrMetProArgLeuAs 171
495 CAATTGTGTCAGATGCGAAGACCTGTGAG...ATGACAGATTATATG 541

```

```
171 pLyspHevalThrcLlValAlaIroIleGlAlaIseThrAlaSerSera 188
542 ATGGGTGGAATACGGCTGATTTAATTAATATACCGATCGCTGTCCAAATC 591
188 sprLaGlyThrTyraSnsPglInsLysTyrrProAlaPheValArgLeu 204
592 GGAAGCAGCAGACAAATATTTGGCGGTCTGATGAAGACGA..... 630
205 GlySerGlySerGlnPheIleTyrrLysGlyAspSntyrSerLeuI 221
631 .CCCAATACCGCGGAAAGTTCAATAT.....ATTGCAAGC 667
221 eLeuAsnAsnHisLcLValAlcLgLYAsnSndLeuLysLeuValGlyAsp 238
668 CATATTCCTGGCTCGTGGTGAATACCTTTGCACAAATAGATGATGAGT 717
238 leryThrTyrrGlyLleAlaGlyThrProTyrrLysValAsnHisGluAsn 254
718 GGTGGCAGACGTCAACTTAGGTAGCAGAAATAATTAACATAGC...CCATA 764
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGlnHisSerAspProLy 271
765 TGCTTTTACCAACAGCA.....GGCTATTGGCGACA 799
271 scLylIleLeuSerGlnAspProLeuThraSntyrAlaValLleuGlyAspS 288
800 GTGGCTCACCAAATGTTTATCTATGATGCCAAAGCAAAAGTGTTAAAT 849
288 ecrLyserrProLeuPheValTyraSparGlnLysGlyLysStrPheupe 304
850 AATGGGTATTGCMAACAGCAACCCCTATATAGGAAAGCAATGGCTT 899
305 LeuGlySerTyraSpherTrpAlaGlyTyraSndLysLysSerTrpGlnI 321
900 CCAGCTAGTTCGTAAGATTGGTCTATGATGAATCTTGTCTGAGATA 949
321 utrpaSntIeTyrrLysProGluPheAlaLysThrValLeuAspLysAsp 338
950 CCCATTCACTATTCAGAACCCATCAAAATGCGAAATACTTTTAAAC 999
338 hr..... 338
1000 GACATTAATATGGCGGAGGAAATCGATGCCAAACATTAACACTATTC 1049
339 .....AlaGlySerLeuThrGlySerAsnThrGlnTyraS 350
1050 TCTACCTTATAGATTAAACACGAACCGCT.....CAATGT 1087
350 utrpaSnpProThrGlyLysThrSerValIleSerSndLysSerGluSerL 367
1088 TTAATGTTCTTTATCCGAGACAGACAGAAACCTGTTATCATGCTGCA 1137
367 euAsnValaLAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGGTCAACAGTTATCGACCCAGACTGAATATATGAGAAATATTC 1187
378 .....ThrAspSerLysLysAsnHisGlyLysSerValTh 390
1188 CTTTATTCAGAAAGGAAAGTGAATGATCTACCGCAACATCAAC 1237
390 rLeu.....ArgLysSerGlyThrLeuThrLeuAsnHisGlnIleAspG 405
1238 AAGGCGCGGCGGTTGTATTTTGAGGTAATTTTACGCTC...TCGCGCT 1284
405 lngLylleGlyLcLLeuPhePheGlnGlyAspTyrrGlyValLysGlyThr 421
1285 AAAACAAACGAAACGTGGCAAGCGCGGCTTCATATCAGTGAGTGAG 1334
422 SerAspSerThrThrTrpLysGlyAlaGlyLysValAlaAspGlyLys 438
1335 TACGCTTACTTGGAAAGTAAAGCGCGTGGAACAGACCGCTGTCCAAA 1384
438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLysI 455
1385 TCGCAAAAGCAGCGCTGCTGTTCAAGCCAAAGGGAAGAACCAAGCGCTG 1434
455 lngLylsGlyThrLeuIleValGlnGlyLysGlyLysGlnLysGlySer 471
1435 CTCACGCGGCGGAGCGTAAGTCAATCTATGATCGACGCGGAGCATCA 1484
472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAs 488
1485 AGGCAAAACAAAGCCTTAGTAATCGCTTGCTGGTGGCGAGGAGGA 1534
488 AsnLysValLysAlaPheSerGlnValGlyLleAlaSerLysArgSerT 505
1535 CGGTGCAACTGAAATCCGATATATCATGTTCAACCCGACAAACTATTTG 1584
505 hrValValLeuAsnAspAspLysGlnValaAspProAsnSerLleTyrrPhe 521
1585 GGCCTTCGCGGCGGACGTTTGATTTGAACGGGCACTGCTTGCTGCCA 1634
522 GlyPheAlaGlyGlyArgLeuAspAlaAsnGlnLysAsnLeuThrPheG 538
1635 CCGCATTTCAAAATACCGATGAAGGCGCGATGATTGTCAACCAATCAAG 1684
538 uHisLleArgAsnIleAspAspGlyAlaArgLeuValAsnHisSntHrs 555
1685 ACAAAGAAATCCACGCTTACCATTCAGGCAATTAAGATATTACT..... 1728
555 eLysThrSerThrValThrIleThrGlyLysSerLeuIleThrAspPro 571
1728 ..... 1728
572 AsnThrIleThrProTyrrAsnIleAspAlaProAspGlnAspAsnProTy 588
1728 ..... 1728
588 rAlaPheArgArgLleLysAspGlyGlnLeuTyrrLeuAsnLeuAla 605
1728 ..... 1728
605 sntTyrrThTyrrAlaLeuArgLysGlyAlaSerThrArgSerGluLeu 621
1729 .....ACAAACCGCAATTAAC..... 1743
622 ProLysAsnSerGlyLysSerAsnGlnAsnTrpLeuTyrrMetGlyLysTh 638
1744 .....AACAACTTGATAGCAAAAAAG 1765
638 rSerAspGlnAlaLysArgAsnValMetAsnHisIleAsnAsnGlnArgM 655
1766 AAATTCCTACAAACGCTTGGTGGCGAAGATGCAACCAAAACGAAAC 1815
655 eTanGlnLysPheAsnGlyTyrrPheGlyLcLlGlnGlyLys...LysAsnAsn 670
1816 GGGCGGCTCAATCGAATTAACCAACCGGAAGACCGGATCGCACTTACT 1865
671 GlyAsnLeuAsnValaThrPheLysGlyLysSerGlnGlnAsnArgPheLe 687
1866 GCTTCGCGGCGGAACAATTTAAACGGCAATATCAGCAAAACAGGCA 1915
687 uLeuThrGlyLcLThrAsnLeuAsnGlnLysAspLeuLysValGlyLysGly 704
1916 AACTGTTTTCAGCGGAGACGACACACCGCACTTACATCATTTAAGA 1965
704 hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 720
1966 AGCGGGTGTCAAAATGCAAGT.....ATCCCAACAGGAAATCGCT 2009
721 lIleSerSerThrLysLysAspGlnHisPheAlaGlnAsnGlnValaVal 737
2010 GTGGGACAAACGATTGATGCAACCGCAATTTAAAGGGAAGAACTTCGATA 2059
737 lValGlnAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnV 754
```

2060 TTCAGGCGGACAAAGCGGTGTTCC... CGCAATGTTGCCAAAGTGAA 2106  
:::.....::: ||| |||||.....:::  
754 alThrAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThr 770  
:::.....::: ||| |||||.....:::  
2107 GGCATTGGCATTTAAGCAATCAGCCCAAGCATGTTCCGTGCGCACCC 2156  
:::.....::: ||| |||||.....:::  
771 SerAsnIleThrAlaSerAspAsnAlaIleValHisIleGlyTyr... 785  
:::.....::: ||| |||||.....:::  
2157 GCATCAAAAGCCACACATCTGTACAGCTTGAGCTGAGCGGTCTACAA 2206  
:::.....::: ||| |||||.....:::  
786 .LysAlaGlyAspThrValCysValArgSerAspTyrThrGlyTyrValr 802  
GTTGTACCGGAAAAAACCATTCAGCATTAAGTGGATTGCTCATTTGAGC 2256  
:::.....::: ||| |||||.....:::  
802 hrcYstIhrThrAspLysLeuSer... AspLysAlaLeuAsnSerPheAsn 817  
:::.....::: ||| |||||.....:::  
2257 AAGACCGCATCAGAGCAATGTCCAGCCTTGCCGATCAGCCTCATTTAAA 2306  
:::.....::: ||| |||||.....:::  
818 AlaThrAsnValSerGlyAsnValAsnLeuSer... 828  
:::.....::: ||| |||||.....:::  
2307 TCTCAGAGGACTTGCCACACTCAAGCGCATCTTAGTCAGGCGGAGACA 2356  
:::.....::: ||| |||||.....:::  
828 ..... 828  
2357 CGCACTATACGGTTACGGGCAAGCCCAACCAAGGCAACCTTCAGCCTC 2406  
:::.....::: ||| |||||.....:::  
828 ..... 828  
2407 GTGGGCAATGCCCAAGCAACATTTAATCAAGCCATTAACGGCAACAC 2456  
||| |||||.....::: ||| |||||.....:::  
829 ...GlyAsnAlaAsnPheValIleuGlyLysAlaAsnLeuPheGlyThrII 844  
:::.....::: ||| |||||.....:::  
2457 ATCGGCTTCGGACAATGCTCATTTAATCTAAGCAACAGCCGCTACAAA 2506  
||| |||||.....::: ||| |||||.....:::  
844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn... 857  
:::.....::: ||| |||||.....:::  
2507 ACGGAGTCTACGCTTCCGACAAAGCTAAGCAACGTAAAGCATTTCC 2556  
:::.....::: ||| |||||.....:::  
857 ..... 857  
2557 GCACCTCAAGCGCAATGTCTCCCTAGCCGTAAGCAGATTTCCATTTTGA 2606  
:::.....::: ||| |||||.....:::  
857 ..... 857  
2607 AAACAGCCGCTTTACCGGAAAAATCAGCGCGCAAGATACGGCATTTAC 2656  
:::.....::: ||| |||||.....:::  
857 ..... 857  
2657 ACTTAAAGACAGCAATGAGACGCTGCCGTCGGGACGGAATTAAGGCAT 2706  
||| |||||.....::: ||| |||||.....:::  
858 ..... SerHisThrPheIleuThrGlyAspSerAsnValAsnGln 870  
:::.....::: ||| |||||.....:::  
2707 TTAACCTTTGACAAAGCCCAACCATTAACCTCAATTCGCCCTATCGACAGA 2756  
||| |||||.....::: ||| |||||.....:::  
871 LeuAsnLeuAspLysGlyHisIleHisLeuAsnAla... GlnAsnAs 885  
:::.....::: ||| |||||.....:::  
2757 TCGGCGAGCGCGCAACCGGCAAGTGGCGGAGATGCCGCCGCCGCGCTT 2806  
||| |||||.....::: ||| |||||.....:::  
885 PAlaAsnLysValThrThr... 891  
:::.....::: ||| |||||.....:::  
2807 CGCGCGCTTCCTATTATTCGTTACGGCGGCAACTTGGCAGAAATCCCGT 2856  
:::.....::: ||| |||||.....:::  
891 ..... 891  
2857 TTCAACAGCGTGCAGTAACGGCAATTTGAACGGTCAGGGAACATTTCCG 2906  
:::.....::: ||| |||||.....:::  
892 TyrAsnThrIleThrValAsn... SerLeuSerGlyAsnGlySerPheTyr 907  
:::.....::: ||| |||||.....:::  
2907 CTTTATGTCGGAATCTTCGGCTACCGGAGCGGCAATTTGAAGCTGCGCG 2956  
:::.....::: ||| |||||.....:::  
907 rTyrIleuThrAspLeuSerAsnLysGlnGlyAspLysValValThrL 924  
:::.....::: ||| |||||.....:::  
2957 AAAGTTCGGAAGGCACTTACACCTTGCGCTGCACAAATACCGGCAACGAA 3006  
:::.....::: ||| |||||.....:::

924 ySerAlaThrGlyAsnPheThrIleGlnValAlaAspLysThrGlyGlu 940  
:::.....::: ||| |||||.....::: ||| |||||.....:::  
3007 CCCGTAAGTCTGCAGCAATTCAGCGTAGTCGAGAGAAAGCAACACACCC 3056  
||| |||||.....::: ||| |||||.....:::  
941 ProThrLys... AsnGluLeuThrIleuPheAspAlaSerAsnAlaThr... 955  
:::.....::: ||| |||||.....:::  
3057 GCTGTCCGAAATCTTAATTTCCACCCCTGCAAAACGAAACAGCTGATGCCG 3106  
956 ...ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAspLeuG 971  
:::.....::: ||| |||||.....:::  
3107 GCGCATGCGGTATTCAGCTTATCCGCAAAAGCGGAGCTTCGCTGCAT 3156  
||| |||||.....::: ||| |||||.....:::  
971 yAlaThrPylsTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyr 987  
:::.....::: ||| |||||.....:::  
3157 AATCCG... 3162  
||| |||||.....:::  
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004  
:::.....::: ||| |||||.....:::  
3162 ..... 3162  
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021  
:::.....::: ||| |||||.....:::  
3162 ..... 3162  
1021 LucIuIleAlaArgValGluThrProValProProAlaProAlaThr 1037  
:::.....::: ||| |||||.....:::  
3162 ..... 3162  
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGluGluSerly 1054  
:::.....::: ||| |||||.....:::  
3163 .....GTCAAGAACAGAGCTTTCGACAAACTCGGCAAGCGCG 3202  
||| |||||.....::: ||| |||||.....:::  
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGluAsnG 1071  
:::.....::: ||| |||||.....:::  
3203 GAGAA... 3207  
||| |||||.....:::  
1071 LysIuValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087  
:::.....::: ||| |||||.....:::  
3208 .....ACAGA 3212  
||| |||||.....:::  
1088 AsnGluValAlaGlnSerGlySerGluThrGluThrGlnThrThrGln 1104  
:::.....::: ||| |||||.....:::  
3213 GCGCGCTTCAGCGCA... 3228  
||| |||||.....:::  
1104 uIleIysGluThrAlaLysValGluLysGluGluLysAlaLysValGluL 1121  
:::.....::: ||| |||||.....:::  
3228 ..... 3228  
1121 ysgLugluLysAlaLysValGluLysAspGluIleGlnGluAlaProGln 1137  
:::.....::: ||| |||||.....:::  
3229 .....AAACAGGCAACACTTGCCGCCAAA... 3252  
||| |||||.....::: ||| |||||.....:::  
1138 MetAlaSerGluThrSerProLysGlnAlaLysProAlaProLysGluVa 1154  
:::.....::: ||| |||||.....:::  
3252 ..... 3252  
1154 lSerThrAspThrLysValGluGluThrGlnValGlnAlaGlnProGlnT 1171  
:::.....::: ||| |||||.....:::  
3252 ..... 3252  
1171 hrcGlnSerThrThrValAlaAlaIaGluAlaThrSerProAsnSerLys 1187  
:::.....::: ||| |||||.....:::  
3253 .....CAACAGCGGGAAGAACAAAGCCCAAGCCTTGA 3287  
||| |||||.....::: ||| |||||.....:::  
1188 ProAlaGluGluThrGlnProSerGluLysThrAsnAlaGluProValTh 1204  
:::.....::: ||| |||||.....:::  
3288 CGGCTGATGTCGGCGCGCGCAAT... 3312  
||| |||||.....:::  
1204 rProValValSerLysAsnGlnThrLysuThrThrAspGlnProThrG 1221  
:::.....::: ||| |||||.....:::  
3313 .....GCCACCGGAAAAAGCAGAAACTGTGGCC 3339  
||| |||||.....:::

```

1221 LuArgLysThrAlaLysValGluThrGluThrGlnGluProPro 1237
3340 GAACCGCGCCGCGAGCGC..... 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValG1 1254
3361 .....GGGAAATGCC...GGCATATTCAGCGCG 3388
1254 nProGlnAlaValLeuGluSerGluValProThrValAsnAlaG 1271
3389 AGGAAGAGAAAAACGGGTGCGAGCGGATTAAGAC..... 3423
1271 LuGlnValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287
3424 .....ACGGCCTT 3431
1288 LysGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrAla11 1304
3432 GCGGAAACGCGCGAAGCGAAACCGCGCGCTAC..... 3468
1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAla15 1321
3468 ..... 3468
1321 eThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337
3469 .....ACGCTTCCCGCGCGCGCGCGCGCGCGCG 3500
1338 ValAlaAsnAsnSerGluSerGluProLysSerAlaArg...ArgAr 1353
3501 GGATTTGCGCGCAACGCGAG..... 3519
1353 gSerIleSerGlnProGlnGluThrSerAlaGluGluThrThrAlaAla15 1370
3520 .....CCCCAACCGCA 3531
1370 eThrAspGluThrThrIleAlaAspAsnSerLysArgSerLysProAsn 1386
3532 CCCCAACCGCAG..... 3543
1387 ArgArgSerArgArgSerValArgSerGluProThrValThrAsnGlySe 1403
3544 .....CGGACCTGATCAGCGCTATGCCAATA 3571
1403 AspArgSerThrValAlaLeuArgAspLeuThrSerThrAsnThrAsna 1420
3572 GCGGTTTGAGTGAATTTCCGCC.....ACGCTCAACGAC 3606
1420 LaValIleSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal 1436
3607 GTTTTCGCCGTACAGACGAATTTGACGCGGTG...TTTGGCAAGACG 3653
1437 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnGluG1 1453
3654 CCGCAACGCGCTTTGGACACGCGGATCCGGGACCAAAACATACGCTT 3703
1453 yGlnItyrAsnValItyrValSerAsnThrSerMetAsnGluAsnItyrS 1470
3704 CGCAAGATTTCCGCGCTACGCGCAACAAACGACCTCGCCAAATCGGT 3753
1470 eSerGlnItyrArgItyrPheSerSerLysSerThrGlnThrGlnLeuGly 1486
3754 ATGCGAAAAACCTGCGCAGCGG...CGCGTGGCATCATGTTTTCGA 3800
1487 TyrAspGlnItyrIleSerAsnAsnValGlnLeuGlyValPheThrTy 1503
3801 CAACCGACCGGAAACACCTTCGACGAGGATGGGCAACCGCGCAGCGC 3850
1503 rValArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuA 1520
3851 TTGCCACGCGTGGCGTTTGGGGAATACGACGCGAGGATTCAGACATC 3900
1520 IaGlnValAsnPheTySerLysTytyrAlaAspAsnHisItyrPtyLeu 1536

```

```

3901 GCATTCACCGCGCGCGGCTTTAGTACGGCGACGCTTTCAGCGCAT 3950
1537 GlyIleAspLeuGlyTyrGlyLysPheGlnSerAsnLeuItyrThrAsnH1 1553
3951 CAGAGCGAAATCCGCGCGCGGTGCTGATTCAGGATTCAGCGCAT 4000
1553 sAsnAlaLysPheAlaArgHisIthrAlaGlnPheGlyLeuThrAlaG1YL 1570
4001 ACCGCGCAGGTTTCGCGGATTCGCGATTCGAAACCGCATTCGCGCAGC 4050
1570 ySalPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyValArg 1586
4051 CGCATTTTCGTCGCAAAAGCGGATTCAGGATTCAGCAAAACGTCATTCGC 4100
1587 TyrSerItyrLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLy 1603
4101 CACCGCGCGCTTCGATTCACCGCTACCGCGCGCGCATTAAGCGCAT 4150
1603 sValAsnProIleSerValIlyThrAlaPheAlaGlnValAspLeuSer 1620
4151 ATTCATTCAAACGCGCGCACATTCATTCATTCAGCCTTATTTGAGCCTG 4200
1620 yThrItyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1635
4201 TCCAT...ACCGATGCCGCTTCGCGCAAGTCGCAACGCGCGTCATAC 4247
1636 ArgTyrAspItyrAsnGlnGlySerGlyLysIle....AsnValAsnG1 1650
4248 CGCGCTATTCGCGCGCATTCGCGCAAAACCGCGCGCATTAAGCGCGC 4297
1650 nTyAspPheAlaItyrAsnValGluAsnGlnGlnItyrAsnAlaG1YL 1667
4298 TAAACCGCGAATCAAGGTTTACGCTGTCCTTCACGCTGCGCGCGC 4347
1667 euItyrLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1683
4348 AAGGCGCGCAATTCGAACGCGACGACGCGCGCATCAATTAAGCTA 4397
1684 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1700
4398 C 4398
1700 e 1700
seq_name: swissprot_40:IGA_NEIGO
seq_documentation_block:
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (rel. 10, created)
DT 01-MAR-1989 (rel. 10, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
DE protease).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RX MEDLINE=87115823; PubMed=3027577;
RA Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT IGA protease."
RT Nature 325:458-462(1987).
RL [2]
RN [2]
RP ACTIVE SITE.
RX MEDLINE=90154052; PubMed=2105953;
RA Bachovchin W.W., Plaut A.G., Flentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of IGA1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids."

```

RL J. Biol. Chem. 265:3738-3743(1990).  
 CC -1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.  
 CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at  
 CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
 CC substrates are known.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC  
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X04835; CA28538.1; -.  
 DR PIR: A26039; A26039.  
 DR MEROPS: S06.001; -.  
 DR InterPro: IPR00710; IGA\_S6.  
 DR Pfam: PF02395; IGA1; 1.  
 DR PRINTS: PRO0921; IGASERPTASE.  
 KW Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;  
 KW Transmembrane; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.  
 FT PROPEP 987 1532 HELPER PEPTIDE.  
 FT ACT SITE 278 278 POTENTIAL.  
 FT SITE 986 987 CLEAVAGE (AUTO-).  
 FT SITE 1018 1019 CLEAVAGE (AUTO-).  
 FT SITE 1121 1122 CLEAVAGE (AUTO-).  
 SQ SEQUENCE 1532 AA; 168976 MW; 68PF4112BD22F40D CRC64;

alignment\_scores:  
 Quality: 1565.50 Length: 1738  
 Ratio: 1.757 Gaps: 46  
 Percent Similarity: 51.266 Percent Identity: 27.100

alignment\_block:  
 US-09-303-518D-653 x IGA\_NEIGO ..

Align seg 1/1 to: IGA\_NEIGO from: 1 to: 1532

```

49 AAAACGGCGCGCATCGCTTCGCGCGCTTACTTACCATATGCGCTGC 98
||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysAlaLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18
99 GTTCGGCATTTGCCCCAAGCCCGCGGCGACACACTTATTTCCGATCA 148
||||| ||||| ||||| ||||| ||||| ||||| |||||
18 aTyAlaLeuThrProTyrSerGluAlaLeuAlaLeuAlaArgAspAspValA 35
149 ACTACCAATATCATCGCGCTTCCGCAAAATAAGCAAGTTGCGCATTC 198
||||| ||||| ||||| ||||| ||||| ||||| |||||
35 sPTyGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
199 GGGCGCAAGATATTGAGTTTACACAAAAAGGAGAGTGGTCGGCAA 248
||||| ||||| ||||| ||||| ||||| ||||| |||||
52 GlYAlaThrAspLeuSerValLysAsnLysArgGlyGlnAsnIleGlyAs 68
249 ATCGATGAGCAAGAGCCCGATGATGATTTCTGTGATTCGCGTAC 297
||||| ||||| ||||| ||||| ||||| ||||| |||||
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85
298 ..GGCGTGGCGCATTTGGCGGCGCATATATATGTCAGCGTCGACAT 345
||||| ||||| ||||| ||||| ||||| ||||| |||||
85 yArgIleLeuThrValValAspProGlnTyrAlaValSerValLysHis 101
346 .....AAGCGGCGCTATACCATGTTGA 368

```

```

102 AlaLysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAs 118
369 TTTTGTGGCGGAGCGAGCAATCCGATACGACCGCTTTCTTCCAA 418
||||| ||||| ||||| ||||| ||||| ||||| |||||
118 nAspValAlaAspLysGluAsn.....GluTyrAsnGly 129
419 TTGTGAAAGAAATATATTAAGCAGGACTAAGCCCATCTTAATGCG 468
||||| ||||| ||||| ||||| ||||| ||||| |||||
129 aValaGluGlnAsnAsnTyrGluPro.....HisLysAlaTrp 141
469 GGC.....GATATACATATGCCGCTTTCGA 494
||||| ||||| ||||| ||||| ||||| ||||| |||||
142 GlYAlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAs 158
495 CAAATTTGTCAGATGCGAAGCCTGTAGATGACCACTTTATGATGAG 544
||||| ||||| ||||| ||||| ||||| ||||| |||||
158 nLysPheValThrGluValAlaProThrAspAlaGlyGlyG 175
545 GGTGG...AAATACGCTGATTAATAATATACCGTATGCTGCAATC 591
||||| ||||| ||||| ||||| ||||| ||||| |||||
175 LysLeuAspThrTyrLysAspLysAsnArgPheSerPheValArgIle 191
592 GGAGCAGCGCAGACA.....TATTGGCGTGTGATGAAGCAACCCAA 635
||||| ||||| ||||| ||||| ||||| ||||| |||||
192 GlYAlaGlyArgGlnLeuValTyrGluLysGlyValTyrHisGlnGluG 208
636 TAACCGCGAAATTCATAT.....CATTTGCAACGCAATTTCT 676
||||| ||||| ||||| ||||| ||||| ||||| |||||
208 yAsn...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgT 224
677 GCGCTCGGTGGTGC..... 690
||||| ||||| ||||| ||||| ||||| ||||| |||||
224 yAlaIleLeuAlaGlyThrProTyrLysAspLysAlaSerGlnThrMet 240
691 AATACCTTTGCAAAATGATGATGCTGCGCAGCTGCACTTATGATG 740
||||| ||||| ||||| ||||| ||||| ||||| |||||
241 AsnThrGluGlyLeuIleGlyPheGlyAsnHisAsnLysGlnTyrSerAl 257
741 CGAAAAATTAACATAGC...CCATATGCTTTTACCAACAGGAGCT 787
||||| ||||| ||||| ||||| ||||| ||||| |||||
257 aGluGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyV 274
788 CATTTGCGCAGCTGCTCAGCAATGTTATCATGATGCCCAAGCAA 837
||||| ||||| ||||| ||||| ||||| ||||| |||||
274 alleuGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsn 290
838 AAGTGGTTAATTAATGGGTATTGCAACAGCAACCCCTATATAGAA 887
||||| ||||| ||||| ||||| ||||| ||||| |||||
291 GlnThrValPheLeuGlyThrTyrAspTyrTyrPalaGlyTyrGlyLysL 307
888 AAGCAATGGCTTCAGTAGTGTGTAAGATGTTGTTATGATGAATCT 937
||||| ||||| ||||| ||||| ||||| ||||| |||||
307 sSerTyrPheGlnIleTyrAsnIleTyrLysLysGluPheAlaAspLysIle 323
938 TTGCTGAGATACCCATTCAGTATCTACGACACCATCAAAATGGGAAA 987
||||| ||||| ||||| ||||| ||||| ||||| |||||
324 .....LysGlnHisAspAsn..... 328
988 TACTTTTAAACGACAATATATATGCGCAGGAAATATGATGCCAACA 1037
||||| ||||| ||||| ||||| ||||| ||||| |||||
329 .....AlaGlyThrValLysGlyAsnG 336
336 yGlnHisHis.....TrpLysThrThr..... 343
1088 TTAATGTTCTTATTCGAGACAGCAAGAACCTGTTATCATGCTGCA 1137
||||| ||||| ||||| ||||| ||||| ||||| |||||
343 ..... 343
1138 GGTGGGTCAACAGTTAT.....CGACCCAGACTG..... 1167
||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

344 ...GlyThrAsnSerHisIleGlySerThrAlaValArgLeuAlaAsnAs 359
1168 .....AATAAGAGAAATATTTCTTATTGACAAAGAA 1204
359 nGluGlyAspAlaAsnAsnGlyGlnAsnValThrPheGluAsp.....A 374
1205 AAGGTGAATTGATTAACAGCAACATCAACCAAGCGCGCGGTGG 1254
374 snGlyThrLeuValIleuAsnGlnAsnIleAsnGlnGlyAlaGlyLeu 390
1255 TATTTAGGGTAAATTTAGGCTCGCCCTAAACACAGAA...ACGTG 1301
391 PhePheSerGlyAspPyrThrValIleGlyAlaAsnAsnAspIleThrTr 407
1302 GCAAGCGCGCGGCTCATATCATGATGATGACATACCGTACTTGGAAAG 1351
407 PLeuGlyAlaGlyIleAspValAlaAspGlyLysValValTrrGlnV 424
1352 TAAACGGCGGTGGCAACAGACCGCTGTCCAAATTCGGCAAGCGCATG 1401
424 aLlysAsnProAsnGlyAspArgLeuAlaIleGlyLysGlyThrLeu 440
1402 CTGGTTCAAGCCAAAGGGAACCAAGGCTCGTCAAGCGTGGCGCAGG 1451
441 GluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLysValGlyLysP 457
1452 TAAAGTCATCTTGAATCAGCAGCGGAGCATCAAGGCAAAACACAGCCT 1501
457 yThrValIleLeuAsnGlnIleLysAlaAspLysLysValGlnAlaP 474
1502 TTGAGTAATCGGCTTGGTCAGCGGAGGCGGAGCGGTGCACTGAATGCC 1551
474 hSerGlnValGlyIleValSerGlyArgGlyThrLeuValIleuAsnSer 490
1552 GATATACGTTCAACCCGCAACAACCTATTTGCGCTTGGCGGCGAGG 1601
491 SerAsnGlnIleAsnProAspAsnLeuYrPheGlyPheArgGlyValYr 507
1602 TTGGGATTGGAACGGGCATTCGCTTTCGTTCCACCGCATCAAAATCCG 1651
507 gLeuAspAlaAsnGlyAsnAspLeuThrPheGlnHisIleArgAsnValA 524
1652 ATGAAGGGCGATGATTCACACCAATCAACAGACAAACAAATCCACGCT 1701
524 spGlnGlyAlaArgIleValAsnHisAsnThrAspHisAlaSerThrIle 540
1702 ACCATTACGCAATTAAGATATTACTACACCGGC..... 1737
541 ThrLeuThrGlyLysSerLeuIleThrAsnProAsnSerLeuSerValHi 557
1737 ..... 1737
557 sSerIleGlnAsnAspTyrAspGluAspAspTyrSerTyrTyrArgP 574
1737 ..... 1737
574 roArgArgProIleProGlnGlyLysAspLeuTyrTyrLysAsnTyrArg 590
1737 ..... 1737
591 TyrTyrAlaLeuLysSerGlyLysArgLeuAsnAlaPrometProGluAs 607
1738 .....AATAAC..... 1743
607 nGlyValAlaGluAsnAsnAspTrrIlePheMetGlyTyrThrGlnGln 624
1744 .....AACACCTGGATAGCAAAAGAAATTTGCC 1773
624 LuAlaArgLysAsnAlaMetAsnHisLysAsnAsnArgArgIleGlyAsp 640
1774 TACAACGGTGGTTGGCGAGAAAGATCAACCAAAAGCAAGCGCGGCT 1823
641 PheGlyGlyPhePheAspGlnGluAsnGlyLysGlyHisAsnGlyAlaLe 657

```

---

```

1824 CAATCTGAATTTACCAACCGGAAGACGGAGCTTACTGCTTCCG 1873
657 uAsnLeuAsnPheAsnGlnGlyLysSerAlaGlnLysArgPheLeuThrG 674
1874 GCGGAACAAATTTAAAGCGCATATCAGCAACCAACAGCGCAATCTGTT 1923
674 yGlyAlaAsnLeuAsnGlnGlyLysLeuSerValThrGlnGlyAsnValLeu 690
1924 TTCACGGCGAGACGACACCGCAGCTTACATCATCTTTAGGAAGCGGCT 1973
691 LeuSerGlyArgProThrProHisAlaArgAspPheValAsnLysSerSe 707
1974 CTCAAAATGGAAGT.....ATCCCAAGAGAGAAATCTGTGGGACA 2017
707 rAlaArgLysAspAlaHisPheSerLysAsnAsnIleValIlePheGluA 724
2018 ACGATTGGATGACCGCACATTTAAAGCGGAAACTTCCATATT...CAG 2064
724 spAspTrrIleAsnArgThrPheLysAlaAlaGlnIleAlaValAsnGln 740
2065 GCGGACAAAGCGGTGTTCCCGCAATGTGCCAAAGTGAAGCGGATG 2114
741 SerAlaSerPheSerSerGlyArgAsnValSerAspIleThrAlaAsnIl 757
2115 GCATTTAAGCAATCAGCCCAAGCGCTTTCGGTGTCCACCGCATCAA 2164
757 eThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr....LysAsn 772
2165 GCCACACAATCTGTACAGCTTCGACTGAGACGGGTGACAAAGTTGAC 2214
772 yAspGlyValLysValAlaArgSerAspTyrThrGlyTyrValThrCysAsn 788
2215 GAAAAACCATTTACCGACGATAAGTGATTCCTTCATTTAGCAAGACCGA 2264
789 ThrGlyAsnLeuSer...AspLysAlaLeuAsnSerPheAspAlaThrAr 804
2265 CATCAGACGAATGTCAGCTTGGCGATCAGCGTCATTTAAATCTCAAG 2314
804 gLleAsnGlyAsnValAsnLeuAsnGlnHisAlaLeuValLeu...G 820
2315 GACTTGCCACACTCAACGCGCAATCTTAGTGAGGCGGAGACACGCACTAT 2364
820 yLysAlaAlaLeuThrPheLysIleGlnGly..... 830
2365 ACGTTTACGCGCAACGCCCAACCAAGGCGACCTCGTGGGCAA 2414
831 ..... 831
2415 TGCCCAAGCAATTTAATCAAGCCACATTTAAAGCGACACATCGGCTT 2464
833 nSerArgValSerLeuAsnGln..... 840
2465 CGGACATGCTTCATTTAATCTAAGCAACACGCCGTACAAAGCGCAGT 2514
840 ..... 840
2515 CTGACGCTTCCGACACAGCGTAAGCAACGTAAGCATTCGCGACTCAA 2564
840 ..... 840
2565 CGGCAATGTCTCCTAGCCGTAAGGCGATTTCCATTTGAAAACAGCG 2614
840 ..... 840
2615 GCTTTACCGGAAATATCAGCGCGGCAAGGATAGCGATTACACTTAANA 2664
840 ..... 840
2665 GACAGCAATGAGACGCTGCGCGTGGCGAGCGAATTTAGCAATTTAACT 2714
841 HisSerLysThrPheLeuThrGlyAspSerGlnValHisAsnLeuSerLe 857

```





```

1413 aglylIeIrgTYSeRrArgLeuSerSerIaAspIryILySLeuGIyAsPa 1430
4091 TCAATATGCGCACCCGGGCGCTTGCAATTCACCGCTACCGGGCGGCAAT 4140
      ::::: :::::||||| :::: |::::|
1430 spSerValIySValSerSerMetIaValIyThrLeuThrAlaGlyIeu 1446
      ::::: :::::||||| :::: |::::|
1411 AAGCGAGATTATTCATTCATCAACCGCGCGCACACATTTCCATACGCGCTTA 4190
      ||| |::::| :::: :::::|||||
1447 AspPheIaIaTyRArgPheLys...ValGIyAsnLeuThrValIySProLe 1462
      ||| |::::| :::: :::::|||||
4191 TTTGAGCCCTGCTCCTATACGATGCGCGCTCCGCGAAATCGGAACGGCGCG 4240
      ||::::| ::::| :::: |::::|
1462 uIeSerIaIaIaTyRphe...AlaAsnTyCIyLySLeIyAlaIasnV 1478
      ::::: ::::| :::: |::::|
4241 TCAATTCACCGCGCATATGGCGCGCAGATTTTCGGCAAAACCGCAGTCGGAA 4290
      ::::: ::::| :::: |::::|
1478 alGIyLyIySerPheAlaTyRILySAlaAspSngInGInIySer 1494
      ::::: ::::| :::: |::::|
4291 TGGGGGCTAAACGCGGAATTCAAAGCTTTCACCGCTGTCCTCCACGCTGC 4340
      ||::::| ::::: ||::::|
1495 AlAGIyValAlaLeuLeuTyRArgSsnValThrLeuAsnValIaSnGlySe 1511
      ::::: ::::: ||::::|
4341 CGCGCGGAGGGGCGCAATTTGGAAGCGACACACGCGGGCGGCTCAAT 4390
      :::::||||| |::::| :::: |::::|
1511 rIleThrLySGLySGLnLeuGIyGInLySserCIyGInIleIyS 1528
      :::::||||| |::::| :::: |::::|
4391 TTAGCTACCGCTGG 4404
      :: |::::|
1528 IeGIInIleArgPhe 1532

```

seq\_name: SwissProt\_40:IGAA\_HAEIN

```

seq_documentation_block:
ID      IGA1_HAE1N      STANDARD:      PRT:      1849 AA.
AC      P45386;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      01-NOV-1995 (Rel. 32, Last annotation update)
DE      Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN      IGA1.
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC      Haemophilus.
OX      NCBI_TaxID=727;
      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-NHTI HK61;
RA      MEDLINE=92234949; PubMed=1373717;
RT      Poulsen K., Reinholdt J., Killian M.;
RT      "A comparative genetic study of serologically distinct Haemophilus
RT      influenzae type 1 immunoglobulin A1 proteases.";
RL      J. Bacteriol. 174:2913-2921(1992).
CC      -1- FUNCTION: VIRULENCE FACTOR. CLEAVES HOST IMMUNOGLOBULIN A
CC      -1- PRODUCING INTACT FC AND F&A FRAGMENTS.
CC      -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC      certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC      substrates are known.
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC      SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC      OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC      DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC      SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASES).
-----
CC      THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC      BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC      THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON
CC      USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC      MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC      ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/
CC      OR SEND AN EMAIL TO license@isb-sib.ch).
CC

```

DR	EMBL: M87491: AAA24966.1.-.	
DR	MEROPS: S06.001: -..	
DR	InterPro: IPR000710: IGA_S6.	
DR	Pfam: PF02395: IGA1: 1	
DR	PRINTS: PRO0921: IGASERPTASE.	
KW	Hydrolase; Serine protease; Transmembrane; Zymogen; Signal	
FT	SIGNAL	1-23
FT	CHAIN	26-1031
FT	PROPEP	1022-1849
FT	ACT_SITE	299-299
SQ	SEQUENCE	1849 AA: 202957 MW: 7947D108C715DAEA CR664;

```
alignment_scores:
  Quality: 1516.50
  Ratio: 1.523
  Percent Similarity: 51.956
  Length: 1917
  Gaps: 60
  Percent Identity: 25.456
```

alignment\_block:

US-09-303-518D-653 X IGA4\_HAEIN

Align seg 1/1 to: IGA4\_HAEIN from: 1 to: 18499

```
205 LeuGlySerGlySerGlnPheIleTyrLysGlySerArgTyrGlnLeu 221
610 .TGGCGGTCTGATGAGACGAACCCATTAACCGCGGAAGTATATCAT. 657
221 uIleLeuThrlGlnLysAspLysGlnGlnLysnLeuLeuArgAsnTrpAspV 238
658 .....ATTGCAAGCGCATATTCCTGGCTGCTC 684
238 aLgLyIAspAsnLeuGlnLeuValGlyAsnAlaTyrThrTyrGlyIle 254
685 GGGGCAATACCTTTCACAAATGGATCGAGTGGTGGCAAGTCAACAT 734
255 AlagLyIhrProTyrLysValAsnHisGlnAsnGlnLysnLeuIleGlyPh 271
735 AGGTAGCAAAAAATTAAACATAGC...CCATATGGTTTTTTTACCAACAG 781
271 eGlyAsnSerLysGlnGlnLysSerAspProLysGlyIleLeuSerGlnA 288
782 GA.....GGCTCATTTGGCGACAGTGGCTCACCAGATGTT 816
288 sPProLeuThrAsnTyrAlaValIleuLysPserGlySerProLeuPhe 304
817 ATCATATGATGCCAAAGCAAAAGTGGTATTAATGGGGTATTCGAAC 866
305 ValTyrAspArgGlnLysGlyLysTrpLeuPheLysGlySerTyrAspPh 321
867 AGGCACCCCTATATAGAAAAAGCAATGGCTTCACACTGTCGTAAAG 916
321 eTrpAlaGlyTyr.....AsnLysLysS 329
917 ATTGGTCTATGATGAATCTTTCGTGAGATACCATTCAGTATTCATC 966
329 eTrpGlnGlnTrpAsnLetyrLysHisGlnPheAlaGlnLysIleTyr 345
967 GAACCAATCAAAATGGGAATACTTTTAAACACAAATAAATAGCGGC 1016
346 GlnGlnTyr.....SerAl 350
1017 AGGAAAAATCATGCAACATTAACATATCTCTACCTATATGATTA 1066
350 aGlySerLeuThrGlySerAsnThrGlnTyrThrTrpGlnAlaThrGlyS 367
1067 AAACACGAACCGTT.....CAATGTTTAAATGTTCTTTATCC 1104
367 eTrhSerThrIleThrGlyGlyGlnProLeuSerValAspLeuThr 383
1105 GAGACAGCAAGAACCTGTTTATCATGCTCAGGTGGGTCAACAGTTA 1154
384 AspGlyLysAspLysPro..... 389
1155 TCGACCCAGACTGAATTAATGAGAAATATTTCTTATTCACAAGGAA 1204
390 .....AsnHisGlyLysSerIleThrLeu.....LysGlyS 400
1205 AAGGTGAATGATACTACACGAACATCAACCAAGCGCGGGGGTGTG 1254
400 eArgLyIhrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyLeu 416
1255 TATTTTGAAGGTAAATTTACGTC...TCGGCTAAAAACAAGCAAGAACTG 1301
417 PhePheGlnGlyAspTyrGlnValLysGlyThrSerAspSerThrThrT 433
1302 GCAAGCGCGGGCTTCATATCAGTATGAGTACGATACCGTCTACTTCGAAG 1351
433 PylsGlyAlaGlyAlaSerValAlaAspGlyLysThrValThrTrpLysV 450
1352 TAAACGGCGTGGCAACGACCGCTGTCCAAATGCGCAAGGACGCTG 1401
450 aHisAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeu 466
1402 CTGGTTCAAGCCAAAGCGGAAACCAAGGCTCGTCAAGCGGGGAGCG 1451
467 ValValGlnGlyLysGlyLysAsnGlnGlyLeuLeuLysValGlyAspGl 483

1452 TAAAGTACTTTAGATACAGCGCGGACGATCAAGCAAAAAACAAGCT 1501
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnLysValGlnIleAsp 500
1502 TTAGTGAATCGGGCTGTGTACGCGGAGGAGCGATGCAATGACC 1551
500 neserGlnValGlyIleValSerGlyArgSerThrIleValLeuAsnAsp 516
1552 GATAATCAAGTTCACCCCGCAAACTATTTGGGCTTTGGCGGGGAGC 1601
517 AspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyArg 533
1602 TTTGGATTTGAAGCGGCTTCGCTTGTTCACCGCTTCAAAATACCG 1651
533 gLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleLeuArgSnIleA 550
1652 ATGAAGGGCGCGATGATGTCAACCAACAAATCAAGCAAGAAATCCACGTT 1701
550 sPAspGlyAlaArgValAlaAsnHisAsnMetThrAsnThrSerAsnIle 566
1702 ACCATTTACAGCAATTAAGATATTACTACA.....ACCGCAA 1739
567 ThrIleThrGlyGlnSerLeuThrAsnProAsnThrIleThrSerTyr 583
1740 TAAAC..... 1743
583 rAsnIleGlnAlaGlnAspAspAspHisProLeuArgIleArgSerIleP 600
1743 ..... 1743
600 rOTyrArgGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeu 616
1743 ..... 1743
617 LysLysGlyAlaSerThrArgSerGlnLeuProGlnAsnSerGlyLys 633
1743 ..... 1743
633 rAsnGlnAsnTyrPleuTyrMetGlyArgThrSerAspGlnAlaLysArgA 650
1744 .....AACAACTGGATAGCAAAAAAATGGCTTCAACAGCTGG 1785
650 snValMetAsnHisIleAsnAsnGlnArgMetAsnLysPheAsnGlyTyr 666
1786 TTTGGCGAGAAAGATGCAACCAAAACG...AACGGCGGCTCAATCTGAA 1832
667 PheGlyGlnGlnGlnLysThrLysAlaThrGlnAsnGlyLysLeuAsnValTrh 683
1833 TTTACCAACCGGAAGAAGCGGATCGCACTTACTGCTTCCGGCGGAACA 1882
683 rPheAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrA 700
1883 ATTTAAAGGCAATATACGCAAAACCAAGCGCAAACTGTTTTCACGGC 1932
700 snLeuAsnGlnLysPheLeuAsnValGlnLysGlyThrLeuPheLeuSerGly 716
1933 AGACCGACCGGACCGCTTCAATCATTTAAGAACGGGCTGTCACAAAT 1982
717 ArgProThrProHisAlaArgAspIleValGlyIleSerSerThrLys 733
1983 GGAAGGT.....ATCCCAACAGAGAAATCGTGGGACCAACGATGGA 2026
733 sAspProHisPheThrGlnAsnAsnGlnValValAlaGlnAspAspTrpI 750
2027 TCGACCGCACATTTAAAGCGGAAACCTTCATATTCAGCGCGGACAGCG 2076
750 lAsnAlaArgAsnPheLysAlaThrThrMetAsnValThrGlnAsnAlaSer 766
2077 GTGGTTTCC...CGCAATGTTCACAAAGTGAAGGCGATTTGGCATTAAG 2123
767 LeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSe 783
```

```

2124 CAATCAGCCCAAGCAGTTTCGTGTCGACCGCATCAAGCCACACA 2173
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
783 rAsnAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspTrnV 798
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2174 TCGTACAGCTTCGGACTGGAGGCTCTACAGATTGACGAAAAAAC 2223
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
798 aLcysValArgSerAspTyrTrnGlyValThrCysHisAsnSerAsn 814
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2224 ATTACGACGATTAAGTGTATGTCATTGACAGACAGACATCAGAG 2273
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
815 Leuser...GluLysAlaLeuAsnSerPheAsnProTrnAsnLeuArgI 830
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2274 CAATGTCAAGCTTGGCGATCAGCTCATTTAAATCTCAAGAGACTTGC 2323
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 yAsnValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAla 846
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2324 CACTCAAGCGCAATCTAGTGCAGCGCA..... 2352
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 snLeuPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLys 862
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2353 ...GACAGCACTATACGTTACGGCAAGCCAGCC..... 2385
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
863 GluAsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAs 879
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2386 ...CAAAAGCGCAACTCAGCTCGTGGCAATGCCCA..... 2421
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
879 nLeuThrAsnGlyHisIleHisLeu.....AsnAlaGlnAsnAspAla 894
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2422 .....GCACATTTAATCAAGCCACATTTAAAGCGCACATCGGCT 2463
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
894 snLysValAlaThrTyrAsnThrLeuThrValAsn.....SerLeu 907
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2464 TCGGACAAATGCTTCATT.....AATCTAAG 2489
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
908 SerGlyAsnGlySerPheTyrTyrValAspPheThrAsnAsnLysSe 924
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2490 CACACAGCCCGTACAAAGCGCAGT..... 2514
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
924 rAsnLysValValAlaAsnLysSerAlaThrGlyAsnPheThrLeuGlnV 941
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2515 .....CTGACCGTTTCCGAC 2529
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 aLAlaAspLysThrGlyLupProAsnHisAsnGluLeuThrLeuPheAsp 957
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2530 AACCGTAAGCA.....AACGTAAAGCATTCGCGCATCAAGCGCA 2570
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
958 AlaSerAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySe 974
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2571 TGTCTCCCTAGCCGATTAAGCAGTATTCATTGAAAAACAGCGCTTGA 2620
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
974 rVal.....AspArgGlyAlaTrpLysTyrLysLeuArgAsnValA 988
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2621 CCGGAAAATATCAGCGCGCAAGATTAAGCATTAACCTTAAAGACAGC 2670
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
988 snGlyArgTyrAsp.....LeuTyrAsnPro 996
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2671 GAATGACGCTGCCGTGCGCAGCAATTAAGCAATTTAACTTGACAA 2720
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
997 Glu.....ValGluLysArgAsnGlnThrValAspTh 1007
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2721 CGCAGCATTAAGCTATCCGCTATCGACAGATGCGGAGCGCGC 2770
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 rThrAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaG 1024
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2771 AAACCGCAGT..... 2781
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1024 InsrAsnAsnGlnGluIleAlaArgValGluThrProValProProPro 1040
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2782 .....GCGGACATGCGCGCGCGCGCT. 2805
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1041 AlaProAlaThrGluSerAlaIleAlaSerGluGlnProGluThrArgPr 1057
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2805 ..... 2805

```

```

1057 cAlaGluThrAlaGlnProAlaMetGluGluThrAsnThrAlaAsnSer 1074
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2806 .....TCG 2808
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1074 hrGluThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsn 1090
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2809 CGCGCTCCCATATATCCGTTACG..... 2832
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1091 SerGluSerValProSerGluThrThrGluLysValAlaGluAsnProPr 1107
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2832 ..... 2832
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1107 oGlnGluAsnGluThrValAlaLysAsnGluGlnGluAlaThrGluPro 1124
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2833 .....CGCCCACTTCGGCA 2847
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1124 hrProGlnAsnGlyGluValAlaLysGluAsnGlnProThrValGluAla 1140
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2848 GAATCCGCTTCAACAGCGCTCAGCGTTAAC...GCCAAATTGAACGTCA 2894
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1141 AsnThrGlnThrAsnGluAlaThrGlnSerGluGlyLysThrGluGluTh 1157
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2895 GCGACATTCGCTTATGTCGAA..... 2919
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1157 rGlnThrAlaGluThrLysSerGluProThrGluSerValThrValSerG 1174
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2920 .....CTCTTC 2925
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1174 LuAsnGlnProGluLysThrValSerGlnSerThrGluAspLysValVal 1190
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2926 GCCTACCGCAGCGGCAATTGAAGCTG..... 2952
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1191 ValGluLysGluGluLysAlaLysValGluThrGluGluThrGlnLysAl 1207
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2952 ..... 2952
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1207 aProGlnValThrSerLysGluProProLysGlnAlaGluProAlaProG 1224
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2953 .....CGGAAAGTTCGAAAGC 2970
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1224 LuGluValProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnGln 1240
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2971 ACTTACACCTTGCTGTC..... 2988
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1241 ThrGlnProThrThrValAlaAlaAlaGluThrThrSerProAsnSerLy 1257
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2988 ..... 2988
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1257 sProAlaGluGluThrGlnGlnProSerGluLysThrAsnAlaGluProV 1274
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2989 .....AACATACCGGCAAGACAGACCCGTAAGTCTCGAG 3021
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1274 aThrProValValSerGluAsnThrAlaThrGlnProThrGluThrGlu 1290
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3022 CAATTGACGCTAGTGAAGAAAGACAAACACACCCCTGCGCAAAATCT 3071
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1291 GluThrAlaLysValGluLysGluLysThrGlnGluValProGlnVal. 1306
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3072 TAAATTCACCTTCGCAACAGACAGCAGTGCAGCGCGGCGGCTATATC 3121
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 .....AlaSerGlnGluSerProLysGlnGluGlnProAl 1318
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3122 AGCTATTCGCAAGAGCGGAGTCCGCTCGATATACCGGTCAAGAA 3171
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1318 aAlaLysProGlnAlaGlnThrLysProGlnAlaGluProAlaArgGlu 1334
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3172 CAAGAGCTTTCGACAA..... 3189
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1335 AsnValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnPr 1351
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3190 .....CTGCGCAGCGCGGAGAGAAACAGAGCC...G 3217
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1351 oGlnThrGlnSerThrAlaValProThrThrGlyLutThrAlaAlaAsnS 1368
13218 CCTTGACGCAAAACAGCAGCACTGGCCGCAACACAGCGGGA... 3264
::: ||| ::| ||||| |||::| |||::| |||
1368 eTlySerProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrLutPro 1384
3265 ..... AAAGCAACGCGCAAAAGCCT 3284
|||::| ::|::| ::|::| ::|::|
1385 AlaArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnTh 1401
3285 TGACGGCGCATTTGGCGCGCGGCAATGCCACGAAAGCAAGAAATG 3334
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1401 rSerAlaThrValSerThrGlnGlnProAlaLysGluThrSerSerAsnV 1418
3335 TTGCGCAACGCGC... 3348
|| ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1418 aLgLucInProAlaProGluAsnSerLLeAsnThrGlySerAlaThrThr 1434
3348 ..... 3348
1435 MetThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValTh 1451
3349 ..... CGCAG... GCAGCGGCGGAAATG 3370
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1451 rGluAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValA 1468
3371 CGCGCATTTATGACGGCGGAGAGAAAGAAAGG...GTG 3408
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1468 lAsnAsnSerGluSerSerGluSerLysSerArgArgArgSerVal 1484
3409 CAGCGCGATTAAGACACCGCCTTGGCGAAGACGCGAAGCGGAACCG 3458
::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1485 SerGlnProLysGluThrSerAlaGluLutThrThrValAlaSerThrG1 1501
3459 GCCGCGTACC... ACCGCTTCCCGCGCGCGCGC 3490
: ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1501 nGluThrThrValAspAsnSerValSerThrProLysProArgSerValA 1518
3491 GCGCGCGCGGATTTGCCGCAACGCGACCGCAACCGCAACCGCGACG 3540
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1518 rGThrArgArgSerValGlnThrAsnSerGluProValGluLeuPro 1534
3541 CAG... 3543
1535 ThrGluAsnAlaGluAsnAlaGluAsnValGlnSerGluAsnAsnValAl 1551
3544 ..... CGGACCTGATCAGCCGTTATGCCATPACG 3574
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1551 AsnSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaV 1568
3575 GTTTGAGTGAATTTTCGCGC...ACGCTCAACAGCGTT 3609
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1568 alLeuSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGly 1584
3610 TTCGGCTACAGAGCAATGTGACCGCGG...TTTGGCGAAGACCGCG 3656
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1585 LysAlaValSerGlnHisLLeSerGlnLeuGluMetAsnAsnGluGly 1601
3657 CAACGCGCTTTGACAGACGCGCATCCGGACACCAACACTACGCTTGC 3706
: |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1601 nTyrAsnValTrpLLeSerAsnThrSerMetAsnLysAsnTyrSerSerG 1618
3707 AAGATTTCCGCGCTACGCGCAACCAACGCACTCGCAATCGGTATG 3756
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1618 LglnLtyrArgArgPheSerSerLysSerThrGlnThrGlnLeuLtyrP 1634
3757 CAGAAAACCTCGCAGCGGG...CGCGTGGCATCTGTTTGGCACAA 3803
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1635 AspGlnThrLLeSerAsnAsnValGlnLeuGlyValPheThrTyrVa 1651
3804 CCGGACCGGAACACCTTCGACGACGCGCATCGCAACTCGACGCGCTG 3853
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1651 lArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuAlaG 1668

```

```

3854 CCCAGGTCGCGTTTTCGGCAATACGCATCGCAGATTCGACATCGC 3903
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1668 lNValAsnPheTyrSerLysTyrTrpAlaAspAsnHisTrpTyrLeuGly 1684
3904 ATCAGCGCGGCGCGGTTTTCAGCGGACGCTTTCAGACGCGCATCAG 3953
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1685 lLeAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnAsn 1701
3954 AGCGAAATTCGCGCGCGGTCGTCATTCAGCGATTCAGCAAGATAC 4003
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1701 nAlaLysPheAlaArgHisThrAlaGlnLLeGlyLeuThrAlaGlyLysA 1718
4004 GCGCAGTTTCGCGGATTCGCGATTCGACACGACATCGCGGACGCGC 4053
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1718 lAraPheAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTyr 1734
4054 TATTTCGTCGCAAAAGCGGATTCGATACGAAAGCTCAATTCGCGAC 4103
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1735 SerTyrLeuSerAsnAlaAspPheAlaLeuAlaAspArgTyleysVa 1751
4104 CCCGGCCTTGCAATTCACCGCTACCGCGCGGCAATTAAGCATTAAT 4153
: ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1751 lAsnProLLeSerValLysThrAlaPheAlaGlnValAspLeuSerTyrT 1768
4154 CATTCAAAACGCGCGCAACACATTTCCATCAGCGCTTATTTAGCCTGTC 4203
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1768 nTyrHisLeuGlyLgl...PheSerLLeThrProLLeuSerAlaArg 1783
4204 TATTCGATCGCGCTTCGCGCAAGTCGCAACCGCTCAATACCGCCGT 4253
||| |||::|::|::|::|::|::|::|::|::|::|::|::|::|
1784 Tyr...AspAlaAsnGlnGlyAsnGlyLyslLeAsnValSerValTyrAs 1799
4254 ATTGGCGCAGGATTTCCGCAAAACCCGACGTGCGAATGGCGGTAACG 4303
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1799 pPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeuLys 1816
4304 CCGAATTCGAAGTTTTCAGCGTCTCCCTCAGCGCTCGCGCGCAAGGG 4353
::: |||::|::|::|::|::|::|::|::|::|::|::|::|::|
1816 eulysTyrHisAsnValLysLeuSerLeuLLeGlyLysLeuThrLysAla 1832
4354 CCGCAATTCGAAGCGCGACGACGCGCGGCAATTAAGGCTAC 4398
||| ||| |||::|::|::|::|::|::|::|::|::|::|::|::|
1833 LysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1847

seq_name: SwissProt_40:YpJA_ECOLI
seq_documentation_block:
ID YpJA_ECOLI STRAND: PRT: 1569 AA.
AC P52143; P76610; P77017; P77019;
AT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein YpJA.
GN YpJA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NC NCB1_TaxID=562;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12."
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;

```

Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubdaram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiiuchi T.,  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-60.8 min on the linkage map and  
RT analysis of its sequence features.";  
RL DNA Res. 4:91-113(1997).  
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC -1- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
-----  
DR EMBL: U36840; AAA79815.1; ALT SEQ.  
DR EMBL: AE000350; AAC75695.1; -.  
DR EMBL: D90889; BAA16514.1; ALT INIT.  
DR EMBL: D90890; BAA16518.1; ALT INIT.  
DR Ecogene: EG13213; ypja.  
KM Hypothetical protein; Outer membrane; Complete proteome.  
SO SEQUENCE 1659 AA: 162774 MW: 0D006A9A3154596 CRC64;

```
alignment_scores:
  Quality: 359.00
  Ratio: 0.481
  Percent Similarity: 44.306
  Length: 1686
  Gaps: 80
  Percent Identity: 19.395
```

alignment\_block:

US-09-303-518D-653 x YPJA\_ECOLI . .

Align seg 1/1 to: YPJ4\_ECOLI from: 1 to: 1569

```

445 GGGATTACGGCAGCATCCCTTATGGGGGAGTATATCATATGCCCGTTGGCA 490
|||||.....:::.....:::.....:::.....:::
97 G|Y|T|H|T|H|A|N|A|S|H|T|H|I|L|E|I|L|E|S|H|I|S|I|L|Y|G|I|N|A|S|N|V|A|I|H|I 113
445 CAAATTTGTCACAGATGCGAGAACCTGTTGAGATGACCGATATATGATGATG 544
|||||.....:::.....:::.....:::.....:::
113 S|G|Y|G|Y|A|L|S|E|R|S|N|G|I|S|E|T|L|U|I|E|G|I|U|S|E|G|Y|G|Y|T|C|I|N|A|S|P|I 130
545 GGTGGAATATACGGTATTAATAATACCTGATTCGTTCGATCATGGA 594
|||||.....:::.....:::.....:::.....:::
130 I|E|G|Y|S|E|H|I|S|A|S|H|S|H|E|V|A|L|Y|G|I|N|A|A|S|A|N|T|H|T|H|I|E|A|S|H 146
595 CGAGGCGACACATATTTGGCGGCTGTGATGAGACGACCAACCAATTAACCGCA 644
:::|||||.....:::.....:::.....:::.....:::
147 G|Y|G|Y|A|T|G|A|S|I|N|.....:::.....:::.....:::.....::: 150
645 AAGTTCATATCATATTTGCA.....ACGCGATATTCCTGGCTGTGGTG 688
|||||.....:::.....:::.....:::.....:::
151 .....S|E|T|L|H|I|S|P|E|L|S|P|E|L|Y|G|I|L|E|S|E|T|H|T|H|T|H|I|E|G|I|U|S|E|R|G 166
689 GCAATACCTTTGCGACCAAAATGCA.....TCAGGTGGTGCGACAGTGCAAC 732
|||||.....:::.....:::.....:::.....:::
166 L|Y|A|S|N|G|I|N|A|S|P|A|L|Y|L|S|Y|G|I|L|E|S|E|R|S|N|G|Y|T|H|T|H|I|E|L|Y|S 182
733 TTAGGTAGCGCAAAAATTTAA...CATAGCCCATATGGTTTATACCAAC 779
|||||.....:::.....:::.....:::.....:::
183 G|Y|G|Y|A|L|S|E|T|A|S|E|T|A|S|Y|A|G|I|U|G|Y|G|I|S|E|T|A|S|N|G|Y|L|E|U|I|E|A|S 199
780 AGAGAGGCTCA.....T 790
|||||.....:::.....:::.....:::.....:::
199 P|G|Y|G|Y|S|E|R|C|I|N|L|E|V|A|L|Y|S|V|A|L|G|I|N|G|Y|H|S|A|L|A|S|P|G|Y|T|H|T|H|T 216
791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGATGCCCAAAAGCAAAAG 840

```

216 LeasnllyserGlySerGlnAspValValIleInlySerIleuAlaThrAsn 232  
:::|||||:::|:::|:::|  
841 TGGTTAATTAATGGGCGATTATGCAACAGCGACCCCTATATAGA.... 885  
|||||:::|:::|:::|  
233 ThrThrIleasnGly.....GlyArgGlnTyValGluGlnIse 245  
.....|:::|:::|:::|  
886 .....AAAGCAATGGCTTCACAGCTAGTCGA 913  
|||||:::|:::|:::|  
245 rThValGluThrThrThrIleIlyasnGlyGluGlnAlaValTyrg 262  
rThValGluThrThrThrIleIlyasnGlyGluGlnAlaValTyrg 262  
:::|:::|:::|:::|:::|  
914 AAGATTGGTTCATGATGAATCTTCGCGAGATACCCATTCAGATTC 963  
:::|:::|:::|:::|:::|  
262 IuserArgAlaIleuAspThrThrIleGluGlyIleThrIleSerIleu... 277  
TACCAACCATCAAAATGGGAATACTTTTAAACGACAATAAATATGG 1011  
277 ..... 277  
1014 CGCAGAAATATGATGCCAAACATAAACATCTCTACCTATAGAT 1066  
:::|:::|:::|:::|:::|  
278 AsnSerIlySerThrAlaIlyasnThrIle.....TySerG 291  
1064 TAAACACGACACCGTTCAATTGTTAACTGTTCTTATCCAGACAGCA 1113  
|:::|:::|:::|:::|:::|  
291 IyGlyThrGlnIleValAsp.....AsnThrSerThrSerAspValIle 305  
:::|:::|:::|:::|:::|  
1114 AGACAACCTTTATATCATGCTGCGAGGGGCGTCAACAGTTATCGACCG 1166  
:::|:::|:::|:::|:::|  
306 Glu.....ValIyr.....SerGlyGlyValIleuAspValArg.... 316  
1164 ACTGAATATATGAGAAATATTTCTTTTATGACAAAGCAAAAGTCAAT 1213  
||:::|:::|:::|:::|:::|  
317 .....GlyIyThrAlaThrAsnValThrGlnHisAspGlyAlaI 330  
1214 TGATACTTACCACCAATCAACCAACGAGCGGGCGGTTTGATTTTGG 1266  
|:::|:::|:::|:::|:::|  
330 IeuluyStrAsnThrAsnGlyIyThrValSerIyIlyThrAsnSerGlu 346  
|:::|:::|:::|:::|:::|  
1264 GGTAAATTTTACGGCTCGCCTTAAACAAACGAAAGCTGGCAAGCGCGG 1313  
|||:::|:::|:::|:::|:::|  
347 GlYAlaPheSerIle.....HisAsnHisVal....AlaAspAs 358  
|:::|:::|:::|:::|:::|  
1314 CGTTCAATCAAGATGGCAGTACCGCTTCTTGGAAGTAAACGCCGTGG 1366  
:::|:::|:::|:::|:::|  
358 nValIleuGlnAsnGlyGlyHisIleuAspIleAsnAlaTyrgIlySerA 375  
|:::|:::|:::|:::|:::|  
1364 CAACAGCGCCCTGTCCAAAATCGGCAAGGACAGCGTG....CTGGTT 1407  
|||:::|:::|:::|:::|:::|  
375 IasnIyStrIleIleIlyS..AspIyGlyIyThmetSerValIleuThr 390  
|:::|:::|:::|:::|:::|  
1408 CAAGCCAAAGG.....GAAACCAAGGCTCGGTACGCGT 1442  
|||:::|:::|:::|:::|:::|  
391 AsnAlaIlyAsnAlaAspAlaThrArgIleAspAsnGlyIyValmetAspVa 407  
|:::|:::|:::|:::|:::|  
407 IAlaGlyAsnAlaThrAsnThrIleIleSn.....G 418  
|:::|:::|:::|:::|:::|  
1487 GCAAAACAAAGCCTTTTACGTGAATCGCGTGGTCAGCGG.... 1527  
|||:::|:::|:::|:::|:::|  
418 IyGlyIyThrGlnAsnIleAsnAsnTyrgIleAlaThrGlyIyThrAsnIle 434  
|:::|:::|:::|:::|:::|  
1528 ..AGGGGACGCGTGAACATGATGCCGATATACGTTACACCCGACAA 1574  
|:::|:::|:::|:::|:::|  
435 AsnSerGlyIyThrGlnAsnIleIlySerGlyIyValAlaAspThrThrI 451  
|:::|:::|:::|:::|:::|  
1575 ACNCTATTTTCGGCTTTCG.... 1594  
:::|:::|:::|:::|:::|  
451 eIleSerSerGlySerArgGlnValValGlnIlyAspIyThrAlaIleG 468  
|:::|:::|:::|:::|:::|  
1595 GCGGACGTTTGGATTTGAACGGGCAATTCGCTTTCGCTC.... 1632  
|:::|:::|:::|:::|:::|

468 lysersanllesealagllyserleullevaltyrThrlyglyle 484  
1633 ..CACCCATTCAAAATACGATGAGGGCGATGATTCAC... 1674  
485 AlaHisGlyValasnGlnGlnThrGlySerAlaLeuValAlaAsnThrG1 501  
1675 .....CACAAATCAACAAAGAAATCCACCGTTACATTA 1708  
501 yAlaGlyThrAspIleGlnGlyTyrAsnLysLeuSerHisPheThrIleT 518  
1709 CAGCGAATPAA.....GATTTACT 1728  
518 hGlyGlyGluAlaAsnTyrValAlaLeuGluAsnThrGlyGlnLeuThr 534  
1729 .....ACACCGGCGAATACAACTTGATAGCAAAAAGA 1766  
535 ValValAlaLysThrSerAlaLysAsnThrIleAspThrGlyGly 551  
1767 AATTGCTTACAAACGGTGTGGCGAAGAAAGATCAACAAAGCAAG 1816  
551 sleuileVal.....GlnLysGlnAlaLysThrAspSERT 563  
1817 GCGGCTCAT.....CTG 1830  
563 hArgLeuAsnAsnGlyGlyValLeuGluValGlnAspGlyGlyGluAla 579  
1831 AATTGCAACCGGAGAGCGATGCGACTTACTGCTTCC..... 1872  
580 LysHisValGlnGlnGlnSerGlyAlaLeuIleAlaSerThrThSe 596  
1873 .....GCGGAACAAATTTAAACGCAAT..... 1896  
596 rGlyThrLeuIleGlnGlyThrAsnSerTyrGlyAspAlaPheTyrIleA 613  
1896 ..... 1896  
613 rGAsnSerGluAlaLysAsnValValLeuGluAsnAlaGlySerLeuThr 629  
1897 .....ATCAGCAACAAACGGCA 1916  
630 ValValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGly 646  
1917 ACTGTTTTCACGCGAGACGACACCGCCTACATCATTTAGAA 1966  
646 sMetAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyT 663  
1967 GC..... 1968  
663 hrcGlnThrIleTyrAlaSerAlaThrSerAspLysAlaAsnIleLysGly 679  
1969 .....GGGTGTCAAAAATGGAAGGTATCCACAAAG 2000  
680 GlyLysGlnThrValTyrGlyLeuAlaThrGluAlaAsnIleGlnSerG1 696  
2001 AGAAATCGTGGGACACGATTTGATCGACCGCATTTAAAGCGGAA 2050  
696 yGlnGlnIleValAspGlyLysThrGlnLysThr..... 708  
2051 ACTTCATTTACGAGCGCA..... 2070  
709 .....HisIleAsnGlyLysThrGlnThrValGlnAsnTyrGlyLysAla 723  
2071 .....CAGCGGTGTTCCCGCAATGT 2093  
724 IleAsnThrAspIleValSerGlyLeuGlnGlnIleMetAlaAsnGly 740  
2094 TGCCAA.....GTGGAAGCGATTTGATTTAAGCATTCAGGCC 2134  
740 rAlaGlnGlySerIleIleAsnGlyLysSerGlnValValAsnGlnGlyG 757  
2135 AAGCAGTTTTCGTCGACCGCATCAAGCAACGATCTGTACAGCT 2184  
757 LysLeuAlaGluAsnSerValLeuAsnAspGlyLysThrLeuAspValArg 773

2185 .....TCGAGCTGACGGGTCTG..... 2202  
774 GlnLysGlySerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuVal 790  
2203 .....ACAGTTGTACGAAAAAACCATTTACCGACGATTAAGTATG 2245  
790 lAlaThrThrArgAlaThrArgValThrGlyThrArgAlaAspGlyValA 807  
2246 CTTCATTGAGCAGACCGACATCATGAGCAATGTGCTTCGCGATC 2295  
807 lApeSerIleGlnGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsn... 822  
2296 GCTCATTTAAATCTCAAGGACTTGCACATCAACGGAATCTTAGTC 2345  
823 .....GlyGlyValLeuThrValGlnLysAspThrSerSe 834  
2346 A.....GCGGAGACAGCGCATATACGTTACCG 2374  
834 rAspLysThrGlnValAsnMetGlyGly.....ArgGlnIleValLysT 849  
2375 GCAACGCGACCCAAACGCGACCTCAGCCTGCGCATGCGCA... 2421  
849 hLysAlaThrAlaThrGlyThr...ThrLeuThrGlyGlnGlnIle 864  
2422 .....GCAACATTTATCAAGCCACATTAAAC.....GCGACACATC 2459  
865 ValGlnGlyValAlaAsnGlnThrThrIleAsnAspGlyGlyIleGlnTh 881  
2460 GCGTTCGAGAAATGCTTCATTATATCAACCAACGCGGTACAAAG 2509  
881 rValSerAlaAsnGlyGlu...AlaIleLysThrLysIleAsnGlnGlyG 897  
2510 GCAGCTGACGCTTCGACACAGCTAAGCAAC.....GTAACCAT 2553  
897 LThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValGlnAsn 913  
914 SerGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIleGlnIleSerG1 930  
2604 TGAANAACGCCCTTTACCGGAAAAATCAGCGCGACAGATGAGCAT 2653  
930 yThrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuAlaThrAsn 947  
2654 TACACTTAAAGACAGC..... 2670  
947 etLeuLeuGluAsnGlyGlyAsnLeuValLeuAlaGlyThrGluAla 963  
2671 ...CAATGAGCGTCCGCGGACGAGCAATTTAGCATTTAAACCTTGA 2717  
964 ArgAspSerThrValGlyLysGlyAlaMetGlnAsnLeuGlnAs 980  
2718 CACGCCACCACTTACACTCAATTCGCGCTATCGAC... 2754  
980 pSerAlaThr...LysValAsnSerGlyGlnTyrThrLeuGlyArgS 996  
2755 .....GATCGCGCA 2763  
996 erLysAspGluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnValAla 1012  
2764 GGC.....GGCAAAACGGGACGCGGACATGCGCGCGCG 2801  
1013 GlyGlyThrAlaAlaValTyrAlaGlyThrLeuAlaAspAlaSerValSe 1029  
2802 CCGTTCGCGCGCTTCCTATATTCGTTACGCG... 2835  
1029 rGlyAlaThrGlySerLeuSerLeuMetThrProArgAspAsnValThrp 1046  
2835 ..... 2835  
1046 roValLysLeuGlnGlyAlaValArgIleThrAspSerAlaThrLeuThr 1062

```
2835 ..... 2835
1063 LeuGlyasnGlyValAspThrThrLeuAlaAspLeuThrAlaIaIaSer 1079
2835 ..... 2835
1079 gGlySerValTrpLeuAsnSerAsnAsnSerCysAlaGlyThrSerAsn 1096
2835 ..... 2835
1096 ySerLysTrpArgValAsnSerLeuLeuAsnAspLysValTrpLeu 1112
2836 ..... 2836
1113 SerAlaGlnThrAlaIaIaProAlaThrThrAsnGlyIleTyrAsnThrIle 1129
2868 GAGCGTAACGGCAATTCAGCGTCAGGAGACATCCGCTTATGTCGG 2917
1129 uThrThrAsn...GlnLeuSerIleSerGlyAsnPhenylLeuHisThrA 1145
2918 AACCTTCGGCTACCGCAGCGGCAATTCAGCTGCGGAAAGTTCGAA 2967
1145 snValAlaGlySerArgGlyAspGlnLeuValValAsnAsnAlaThr 1161
2968 GGCACCTTACACCTTGCTGTCAACAATACCGCAACGAAACCGTAAGTCT 3017
1162 GlyAsnPhenylLeuPheValGlnAspThrGlyValSerProGlnSerAs 1178
3018 CGAGCAATTCAGCTAGTGGA...GGAAAGCAACACACCGCTGT 3061
1178 pAspAlaMetThrLeuValLysThrGlyLysAsp... 1190
3062 CCGAAATCTTAATTCACCTGCACAAACGAA...CACGTGATGCC 3105
1191 .....AlaSerPheThrLeuGlyAsnThrGlyLysPheValAspLeu 1204
3106 GGGGCAATGGCGTTATCAGCTATCCGCAAGACGGGAGTTCGGCTGCA 3155
1205 GlyThrTrpGlyLys...ValLeuLysSerAspGlyAsnSer... 1217
3156 TATCCGCTCAAAAGAACAGAGCTTTCGCAAACTCGGCAAGCGGGAG 3205
1217 ..... 1217
3206 AACAGAGGCGCGCTTGACGGCAAAACAGGCAACCTGCGCCAAACAA 3255
1217 ..... 1217
3256 CAGCGGCAAAAGACACGCGCAAGCGCTGACGGCGCTGATTCGGCGGG 3305
1218 .....AsnThrAsnLeuThrAsnAspValLysProA 1228
3306 GCGCAATGCACCGCAAAAGGAGAAAGTTCGCGAACCGCGCGCGGAGG 3355
1228 snProAspPro...IleProAsn... 1234
3356 CAGCGGGGGAATATCCCGCATTCAGCGCGGAGAAAGAAAAACGG 3405
1234 ..... 1234
3406 GTGACGCGGATTAAGACACCGCTTGCGAAACAGCGCGGAAGCGGAAC 3455
1235 .....ProLysP 1237
3456 CCGGCGCGCTACACCGCTTCGCCGCGCGCGCGCGCGCGCGGAGATT 3505
1237 roAspProLysProAspProLysProAspPro...AsnProLysProAsp 1252
3506 TCCCGCAACCGGACCGCAACCGCAACCGCAACCGGACGACCTGATC 3555
1253 ..ProThrProAspProThrProThrProValProGlnLysArgGlnThr 1268
3556 AGCCGTATGCATAGCGGTTTGAGTGAATTTCCGCGCAGCTCAACAG 3605

1269 Pro.....SerThrAlaIaIaValLeuAsnMetAlaIaIaThrProLe 1283
3606 CGTTTTC.....GCCGTACAGAGCAATTCGACCGCGTGT 3640
1283 uValPheAspAlaGlnLeuAsnSerIleArgGlnArgLeuAsnIleMetL 1300
3641 TTGCCGAAGACCGCGCAACCGCGTTTGACACAGCGGACATCCGGACAC 3690
1300 yAlaSerProHisAsnAsnValTrp...GlyAlaThrTyrAsnThr 1315
3691 AACACATCCGTTCCGAGATTTCGCGCTACCGCCCAACACCGACCT 3740
1316 ArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGlnGlnThr... 1330
3741 GCGCCAAATCGGTATGCAGAAACCTCGCAGCGGCG...G 3781
1331 ...LeuThrGlyMetThrValGlyIleAspSerArgAsnAspIleProG 1346
3782 TCGGCATCCTG.....TTTTCGCACACCGGAC 3810
1346 LysGlyIleThrThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIle 1362
3811 GGAAACACCTTCGACGAGCGCATCGGCAACTCGGACGCGCTTGCCACGG 3860
1363 Gly.....PheAspArgGly.....GlyHisGln 1370
3861 TGCGCTTTCGGCAATTCGCGATCGC..... 3888
1370 ySerVal...GlySerTyrSerLeuGlyGlyTyrAlaSerTrpLysIsg 1386
3889 .....AGTTTCGACATC 3900
1386 LysSerGlyPheThrLeuAspLysValValLysLeuAsnAlaArgPheLysSer 1402
3901 GGCATTCAC.....GCGGCGCGCGTTTATGACGCGACGCTT.. 3939
1403 AsnValAlaGlyLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrH 1419
3940 ..TCAGACGCGCATCAGAGCAAAATCCCGCGCGGCTCTGCATTCAGGCA 3988
1419 sSerAsnGlyLeuGlyGlyHis.....I 1427
3989 TTCAGGCAAGATACCGCGCAGGTTTCGCGGATTCGCGCATCGAACCGCAC 4038
1427 LeuGlnThrGlyMetArgPheThrAspLysAsnTrpAsnLeuThrProTyr 1443
4039 ATCGGCGCAACGCGCTATTTCGTCAAAAGCGGATTACCGATACGAAA 4088
1444 AlaSerLeuThrGlyPheThrAlaAspAsnProGlnTyrHisLeuSerAs 1460
4089 C.....GTCAATATCCGCACCCCGCGCTTCGATTACCGCT 4126
1460 nGlyMetLysSerLysSerValAspThrArgSerIle.....T 1473
4127 ACCGC...GCGGCGATTAAAGCAGATTCATTCACAAACCGCGCAACAC 4173
1473 yTrpArgGlnLeuGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGly 1489
4174 ATTTCATTCACGCGCTTATTGAGCTGTCTTATACCGATGCGCTTCGG 4223
1490 MetGlnValGlnProTrpLeuLysAlaIaIaValArgLysGlnPheValAs 1506
4224 CAAGATCGCAAGCGCGCTCAATACCGCGCGGATTGCGCGACGATTTC... 4269
1506 pAspAsnArgValLysValAsnSerAspGlyAsnPheValAsnTyrLeuAs 1523
4270 ..GGCAAAACCGCGAGTGCAGAAATGGGCGCTAAACCGCGCAAAAGGT 4317
1523 ergLysArgArgGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer 1539
4318 TTACGCTGTCCCTCCACAGCTGCCCGCGCGCAAGGCGCGCAATTCGAACG 4367
```



```

1540 ...ThlEusergLYHISLeugLYValGLY.....Ty 1549
RT GCAGCAGCGCGCGGCATCAAA.....TTAGGTCACC 4399
CC :|||||:|||||:|||||:|||||:|||||:|||||:
1549 rserHISserHISaGLYValaGIuserPrIOTrPaSnaLaValaLaGLYVala 1566
4400 GCTCG 4404
1566 snTTP 1567

seq_name: SwissProt_40:YFAL_ECOLI
seq_documentation_block:
ID YFAL_ECOLI STANDARD: PRT: 1250 AA.
AC P45508; P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfal precursor.
GN YFAL OR B2233.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Sato Y., Sivasubraman S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horikoshi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-0V6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.J., Salmon G.P.C.;
RT "The parD-mutant of Escherichia coli also carries a gyrA mutation.
RT The complete sequence of gyrA.";
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;

```

```

RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes.";
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -1- SIMILARITY: TO E COLI YDEK.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE000313; AAC75293.1; -
DR EMBL; D90855; BAA16052.1; ALT_INIT.
DR EMBL; D90854; BAA16050.1; ALT_INIT.
DR EMBL; K02672; -; NOT_ANNOTATED_CDS.
DR EMBL; U30459; AAA74094.1; -
DR EMBL; Y00544; -; NOT_ANNOTATED_CDS.
DR Ecocore: E312850; yfal.
KW Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT FT 28 30
FT CONFLICT 40 40 (IN REF. 3).
FT CONFLICT 65 66
FT CONFLICT 431 431
FT CONFLICT 433 434
FT CONFLICT 478 478
FT CONFLICT 773 773
FT CONFLICT 853 853
FT CONFLICT 923 924
FT CONFLICT 948 994
FT SQ SEQUENCE 1250 AA; 131152 MW; 17F98C05E29FC95 CRC64;

alignment_scores:
Quality: 292.50 Length: 1466
Ratio: 0.447 Gaps: 72
Percent Similarity: 44.679 Percent Identity: 19.986

alignment_block:
US-09-303-518D-653 x YFAL_ECOLI ..
Align seg 1/1 to: YFAL_ECOLI from: 1 to: 1250

637 AACCGCGAAGTTCATATTCAGCGCATATTC..... 675
:::|||||:|||||:|||||:|||||:|||||:|||||:
46 serArgInserLeusergLYleThrGLnAsPTrSerleAlaasp1 62
676 ....TGCGTCGCGTCGCAATACCTTGCGCAAAATGATCGAGTGTG 721
62 yGIITrPLeuValPheSerasp...MetInrAsnAlaSerGIyGLy 78
722 GCACAGTCAACTGATGAGCGAAAAAATTAACATAGCCCATATGTTT 771
78 laVal.....Phe 80
772 TTACCAACAGAGGCTCATTT.....GGCAGACTGGCTC 806
|||||:|||||:|||||:|||||:|||||:|||||:
81 leuInGInGlyAlaGluPheSerLeuPProGluSnsIuThrGLyMe 97
807 ACCAATGTTATCTATGATGCGCCCAAAAGCAAAAGTGTATTAATAGGCG 856
97 tThrLeuPheAlaAsnAsnThr.....ValThrGLy 108
857 TATTCGAACAGCGCACCCCTATATAGAAAAAGCAATGGCTCCACGTA 906
:::|||||:|||||:|||||:|||||:|||||:|||||:
108 luTyraSAsnGlyGLyAlaIlePheAlaLysGluAsnSerThrLeuAsn 124

```

```
907 GTTCGTAAGATGTGTTCTATGATGAATCTTTGCTGAGATACCATC 956
    ||| |||||:||||:|:
125 LeuThr ..... AspValIlePheSerGlyAsnValAlaG1 136
957 AGTATTCTACGACACCATCAAAATGGAAATACCTTTTAAAGACATA 1006
    :||| :|||:||||:|:
136 yGlyTyr..... GlyAlaIleTyrSerSerGlyT 147
1007 ATATGGCGGAGGAAATGATGCCAAATATAACTATTCTTCTACT 1056
    :||| :|||:||||:|:
147 hrisnAspThrGlyAlaValAspLeuArg..... 156
1057 TATAGATTAATAACACGACCGTTCAATGTTTAACTGTTCTTATCCG 1106
    ||| :|||:||||:|:
157 ..... ValThrAsnAlaMetPheArgAsnAsnIleAlaAs 168
1107 GACAGCAGACGAACCTGTTTATCATGCTGAGGTGGGCTCAACAGTTATC 1156
    :|||:||||:|:
168 naSpGlyLys..... GlyGlyAlaIleTyrThr.... 177
1157 GACCCAGACTGAATATGAGAAATATTTCTTATTGACAAAGGAAA 1206
    :|||:||||:|:
178 ..... IleAsnAspValTyrLeuSerAspValIlePheAspAsn 191
1207 GGTGAATTGATCTTACCTACGACCAATCAACGAGCGGCGGCTTG... 1254
    :|||:||||:|:
192 AsnGlnAlaIleTyrThrSerThrSerTyrSerAspGlyAspGlyAlaI 208
1255 ..... TATTGTCG 1261
208 eaSpValIThrAspAsnAsnSerAspSerLysHisProSerGlyTyrThrI 225
1262 AGGGTATATTTTACGCTCTCGCTCAAAACAGCAACACGTGGCAAGCGGG 1311
    ||| |||||:||||:|:
225 LeValAsnAsnThrAlaPheThrAsnAsnThrAlaGlnGlyTyrGlyGly 241
1312 GGGGTCATATCAGTGTGATGCGACGTACCTTACTTGGAAAGTAAAGCGCT 1361
    :|||:||||:|:
242 AlaIleTyr... ThrAsnSerValThrAlaProTyrLeuIleAspIleSe 257
1362 GGCAAAGCGCGCTGTCCAAATGCGCAAGGACGACGCTGCTGCTTCAA 1410
    :|||:||||:|:
257 rValAspAspSerTyrSerGlnAsnGly... GlyValLeuValAspGlu 273
1411 ..... GCCAAAGGCGAAACCAAGGCTCGTCAGC..... 1440
    ||| ||| :|||:|:
273 snAsnSerAlaIleGlyTyrGlyAspGlyProSerSerAlaIleGlyGly 289
1441 ..... GTGGGCGACGGTAA 1454
290 PheMetTyrLeuGlyLeuSerGluValIThrPheAspIleAlaAspGlyL 306
1455 AGTCATCTTAGATCAAGAGCGGAGATCAAGGCAAAAAACAGCTTTA 1504
    |||:||||:|:
306 sThrLeuValIleGlyAsnThrGlnAsnAspGlyAlaValAspSerIleA 323
1505 GTGAATGCGCTTGCTC..... AGCGGAGGCGGACGCTGCACTGAAT 1548
    :|||:||||:|:
323 laGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeuAsn 339
1549 GCGGATTAAT..... 1557
340 AlaAspAsnAsnAspPheThrGlyGluMetGlnIleGluAsnGlyGluVa 356
1558 ..... C 1558
356 IThrLeuGlyLysSerAsnSerLeuMetAsnValGlyAspThrHisCysG 373
1559 AGTTAACCAGCCGACAACTATTTGCGCTTCCGCGGACGCTTGGAAT 1608
    |||:||||:|:
373 InAspAspProGlnAspCysTyr... GlyLeuThrIleGlySerIleAsp 388
1609 ..... TTGAACGGGCAATGCTTTCG..... 1629
389 GlnTyrGlnAsnGlnAlaGlnLeuAsnValGlySerThrGlnGlnThrPh 405
1630 .TTCCAGCCGATTTCAAAATACCGATGAGAGGCGGATGATTTGCAACACA 1678
    ||| :|||:||||:|:
405 eValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu..... A 418
1679 ATCAAGACCAAGATCCACCGTTTACC..... 1704
    ||| :|||:||||:|:
418 snIleAspAlaGlyGlyAsnValIThrValAsnGlnGlySerPheAlaGly 434
1705 ...ATTACAGCAATTAAGATATTACTTCAACCGGCAATTAACAAAC.. 1749
    ||| :|||:||||:|:
435 IleIleGlnGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVa 451
1750 .TTGGATGACAAAGAAATGCTTACACGCGTGGTTGGCGGAAG 1798
    ||| :|||:||||:|:
451 IleuAlaGlyAlaGlnSerMetAlaLeuThrGly..... A 463
1799 ATGCAACCAAAAGAGAGCGGCGCTCAATCTGAATTACCAACCGAAGAA 1848
    ||| :|||:||||:|:
463 spIleValValAspAspGlyAlaVal... LeuSerLeuGlnGlyAspAla 478
1849 GCGGATCGCACTTCTGCTTCCGCGGACAAAT.....TTAAACGG 1892
    ||||| ||| ||| :|||:|:
479 AlaAspLeuThrAlaLeuGlnAspAspProGlnSerIleValLeuAsnGln 495
1893 CAATATCACGCAACCAACGCGCAAACTGTTTTCAGCGGACGACGACAC 1942
    |||:||||:|:
495 yGlyValLeuAspLeuSerAsp..... 502
1943 CGACGCGCTTACAAATCATTTAGAAACGGGCTGTCAAAATGGAAGTATC 1992
    :|||:||||:|:
503 ..... PheSerThrTrpGlnSerGlyThrSerTyrAsnAspGlyLeu 516
1993 ..... CCAGAGGAGAAATGCTGCGGACACAGATTGGAATCGA 2030
    |||:||||:|:
517 GlnValSerGlySerSerGlyThrValIleGlySerGlnAspValValAs 533
2031 CCGGACATTTTAAAGCGGAAATCTTCATATTCACAGGC...GGACAAACGG 2077
    ||| :|||:||||:|:
533 PheAla... GlyGlyAspAsnLeuHisIleGlyGlyAspGlyLysAspG 549
2078 TGGTTTCCCGCAATGTTGCCAAAGTGCAGAGCGCATTTGAAGCAAT 2127
    ||| ||| ||| :|||:|:
549 lValIlyrValValValAspAlaSerAspGlyGlnValSerLeuAlaAsn 565
2128 CAGCGCCAGACGATTTCCGCTGCTGCGACCGCATCAAGCCACACAAATC 2175
    :|||:||||:|:
566 Asn...AsnSerTyrLeuGlyThrThrGlnIleAlaSerGlyThrLeuMe 581
2176 .TGTAACGTTTGGAGCTGAGCGGCTTGACAAAGTTTGCGAAAAAAACA 2224
    :|||:||||:|:
581 tValSerAspAsnSerGlnLeuGlyAspThrHisTyrAsnArgGlnValI 598
2225 TTACCGACGATTA... ..GTGATTCCTTATTGACCAAGACC 2262
    |||:||||:|:
598 lPheThrAspLysGlnGlnGlnSerValMetGluIleThrSerAspVal 614
2263 GACATCAGACGCAATGTCAGC.....CTGC 2288
    ||| :|||:||||:|:
615 AspThrArgSerAspAlaIleGlyHisGlyAlaAspIleGluMetArgAl 631
2289 CGATCAGCGTCATTTAAATCTCACAGAGACTTGCCACACTCAACGCAATC 2338
    ||||| :|||:||||:|:
631 aaSpGlyGlnValAlaValAspAlaGlyValAspThrGlnTrpGlyAla 648
2339 TTATGTCAGCGGGA.....GACACGCAATATACGTTTACG 2373
    ||| ||| :|||:|:
648 eMetAlaAspSerSerGlyGlnHisGlnAspGlnGlySerThrLeuThr 664
2374 CGCAACGCGCAACCAACGCGCAACTGACGCTTCGTCGCAATGCGCAACG 2423
```

```

665 LysThrIleValagIThrLeuGluThrIleValSerGlyThrThrGlnse 661
2424 AMCA.....TTTATCAAGCCACATTAAAGCGCAACATCG..... 2460
681 rAlaValArgValGluGluGlyThrLeuLysGlyAspValAlaAspIlel 698
2461 .....GCTTCGCAATGCTTCATTAACTA 2487
698 euProTyrAlaSerSerLeuTyrValGlyAspGlyAlaThrPheValThr 714
2488 .....AGCAACAAGCCGTACAAAAGCGCACTCT 2516
715 GlyAlaAspGlnAspIleGlnSerIleAspAlaIleSerSerGlyThrI 731
2517 GAGCGTTTCGACAGCCCTAAGCAACGTAGCCATTCCGCACTCAAG 2566
731 eaSpIleSeraspGlyThrValLeuArgLeuThr..... 742
2567 GCATGTCTCCCTAGCCGATAGGCAAGTATTCATTGAAACAGCCGC 2616
742 ..... 742
2617 TTTACCGGAAAAATCAGCGCGGCAAGGATAGCATTTACATTAAAGA 2666
743 .....GlyGlnAspThrSerValAlaLeuAsnAl 752
2667 CAGC.....GAATGAGCGCTGCCGTGGGCAAGGA..... 2697
752 aSerLeuPheAsnGlnLysaspGlyThrLeuValAlaThrAspGlyValT 769
2698 .....TTAGCAATTTAAC..... 2712
769 hrLeuThrGlyGluLeuAsnThrAsnLeuGluThrAspSerLeuThrTyr 785
2713 CTTCGAACAGCCACCATTAACATAATCCGCTATGACACGATGGGCG 2762
786 LeuSerAsnValAlThrValAsnGlyAsnLeuThrAsnThrSerGlyAla 802
2763 AGGCGGCAAAACCGCAGTGGCGGAGATGCGCGCGCGCTGGCGGC 2812
802 lSerLeuGlnAsnGlyAlaAlaGlyAsp..... 811
2813 GTTCCCTATTATCCGTTACGCCGCAACTCTGCGAGAATCCGTTTCAAC 2862
811 ..... 811
2863 ACGGTGAGGTAAACGCAATTTGAAGCGTCAAGGAACATTCGCTTTAT 2912
812 ThrLeuThrValAsnGlyAspTyrThrGlyGlyThrLeuLeuAs 828
2913 GTGCGAAGCTTTGCG.....TACCGCAGCGCAATTTGAAGCTGGCG 2956
828 pSerIleuLeuAsnGlyAspSerValSerAspGlnLeuValMetAsn 845
2957 AAAGTTCGAGGACACTTACACTTGGCTGTCACAAAT...ACGGCG... 3000
845 lYAsnThrAlaGlyAsnThrThrValAlaValAsnSerIleThrGlyIle 861
3001 AACGAACCCGTAAGTCTGAGCAATTGACGTAGTGAAGGAAAAAGACAA 3050
862 GlyGluProThrSerThrGlyIleLysValAlaAspPheAlaIleAsp 878
3051 CACACCGGTGTCGAAAACTTAATTTTCACTGCAAAAAGAA...CAGC 3097
878 oThrGlnPheGlnAsnAsnAlaGlnPheSerLeuAlaGlySerGlyTyr 895
3098 TCGATGCGCGGCGCATGCGCTTATCAGCTTATCCGCAAGAGCGGAGTTC 3147
895 aLAsnMetGlyAlaTyrAspTyrThrLeuValGluAspAsnAsnAsp 911
3148 CGCGTCGATTAATCCGTTCAAGAACAGAGGCTTTCGACAAACTGGGCA 3197

```

---

```

912 TyrLeu..... 913
3198 GCGGAGAGAAACAGAGCGCGCTTACGGCAAAACAGGCAACATTGCCG 3247
913 ..... 913
3248 CCAACACAGCGGAAAAAGACAAACGCGCAAAAGCTTACGCGCTGATT 3297
913 ..... 913
3298 GCGCGCGGCGCAATGCCAACGCAAAAGCAGAAAGTGTGCCGAACGCG 3347
914 .....ArgSerGlnGluValThrProProse 922
3348 CCGGACGCGAGCGGGAAAAATGCCGATATTCAGCGGAGAGAAAGAGA 3397
922 r..... 922
3398 AAAACGGGTGAGCGGATTAAGACACCGGCTTGGCAAAACAGCGGAA 3447
923 .....ProProAspProAsp..... 927
3448 GCGGAACCGCGCGGCTACCAACGCGCTTCCCGCGCGCGCGCGCGCG 3497
928 .....ProThrProAsp..... 931
3498 CCGGATTTGCGCAACCGCACCGCCCAACCGCAACCGCACCGCGCG 3547
932 .....ProAspProThrProAspProAspProThrProAspPro 945
3548 ACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTCCGCCACG 3597
945 lProThrProAlaTyrGlnProValLeuAsnAlaLysAlaIleGlyTyr 961
3598 CTCACAGCGTTTTCGCGCTACAGAGCAAGTTCGACGCGCTTCCGCA 3647
962 LeuAsnAsnLeuArgAlaAla.....AsnGlnAlaPheMetIle 974
3648 AGACCGCGCGCAACCGCGTTTGGCAACGCGC..... 3678
974 tGluArgAspIleAlaGlyLysaspGlyThrLeuAsnLeuArgVal 991
3679 ..ATCGGAGACCAACAACTACCGCTTCGCAAGATTTCGCGCGCTACCG 3726
991 allIleGlyLysPyrThrIleThrAlaAlaGlyIleAla..... 1005
3727 CAACAACCGACCTGCGCCAAATCGGTATGCAGAAAACTCGGCGCG 3776
1006 GlnHisGlnAspThrSerThrValGlnLeuSerGlyAspLeuPheSer 1022
3777 GCGC.....GTGCGCATCCG.....TTT 3796
1022 YArgTyrGlyThrAspGlyLysThrMetLeuGlyIleValGlyGlyTyr 1039
3797 CGCACAACCGAGCGCAACACCTTCGACGAGCGCATCGCAACTCGCA 3846
1039 eAspAsnGln...GlyAspSerArgSerAsnMetThrGlyThrArgAla 1054
3847 ...CGGCTTCGCAACGAGTCCGTTTCGG.....CA 3875
1055 AspAsnGlnAsnHisGlyTyrAlaValAlaGlyLeuThrSerSerThrPhe 1071
3876 ATACGGATGGGACAGTTCGACATCGGCATCGACGCGGCGCGCGGTTA 3925
1071 nHisGlyAsnGlnLysGlnGlyAlaThrPheAspSerThrPheGlnTyr 1088
3926 GT.....ACGCGCAGCGCTTCGACGCGCATCGAGCAAAATCCGCGC 3969
1088 lAtThrPheSerAsnAspValSerGlnGlnLysPhe..... 1100
3970 CCGGTGCTGCATTAC.....GGCATTCAGGCAAGATACCGCGCAG 4010
1101 ...ThrAspHisTyrHisSerSerGlyIleIleAlaSerLeuGluAla 1116

```

```

4011 TTTC.....GGCGATTGCGATGCAACCGACATGCGG 4045
1116 YTYRGINTRPLEUProGLYArgGLYValAlIleGIUProGlnAlaGlnV 1133
4046 CAACGGCGATTTCGTCGCAAAAAGCGATTACCGATGCAAAAC..... 4089
1133 AlIleTYRGINGLYAlGInGlnInAspAspPheThrAlaAlaInAsnArgAla 1149
4090 .....GTCAATATCGCACCCCGCGCGCTTCAT 4118
1150 ArgValSerGlnSerGlnInLysAspAspIleGlnThr..... 1161
4119 CAACCGCTACCGCGCGCATTAAGCAGATATTTCATCAACCGCGCGC 4168
1162 .....ArgLeuGlyLeuHISerGlnThr.....ArgThrAlaV 1173
4169 AACACATTTCATCAGCGCTTATTGAGCGTGTCC...TATACCGATGCC 4215
1173 aHIS.....ValIleProThrLeuAspLeuAsnTYRTrHISAspPro 1187
4216 GCTTCGCGCAAGTCCGACAGCGCGCATACCGCGATTTGGCCAGGA 4265
1188 HISerThrGlnIleGlnLeuAspGlySerThr.....IleSerAspAs 1202
4266 TTTCGGCAAAACCGCG...AGTGGCGAAATGGGCGTAAACCGCGAAATCA 4312
1202 pAlaValLysGlnArgGlyGlnIleLysValGlyAlaThrGlyAsnIleS 1219
4313 AAGGTTTCAGCTGTCCCTCCACGCTGCCCGCGC.....AAGGGCGG 4356
1219 eGln...ArgValSerLeuArgGlySerValAlaIleArgGlnGlySer 1234
4357 CAATTGAGCGCAGACAGCGCGCATCAAAATTAGCGTACCGGCTG 4404
1235 AspAspPheAlaGlnThrAlaGlyPheLeuSerMetThrValLysTrp 1250

seq_name: SwissProt_40:HLVA_PROMI

seq_documentation_block:
ID HLVA_PROMI STANDARD; PRT; 1577 AA.
AC p16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN HPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RX STRAIN-ISOLATE 477-12;
RX MEDLINE=90170827; PubMed=2407716;
RA Uphoff T.S., Welch R.A.;
RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent
RT hemolysin genes (hpma and hpmb) reveals sequence similarity with the
RT Serratia marcescens hemolysin genes (shla and shlb).";
RL J. Bacteriol. 172:1206-1216(1990).
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC DEFINED.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
CC REQUIRES HPMB FUNCTION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA
CC MAY BE RESPONSIBLE FOR PORE FORMATION.
CC -!- SIMILARITY: TO S. MARCESCENS HEMOLYSIN (SHLA).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M30186; AAA25657.1; -
DR PIR: A35140; A35140.
KW Hemolysins; Toxin; Outer membrane; Signal.
FT SIGNAL
FT CHAIN 30 1577
SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

alignment_scores:
Quality: 277.50 Length: 1715
Ratio: 0.373 Gaps: 88
Percent Similarity: 43.382 Percent Identity: 19.767

alignment_block:
US-09-303-518d-653 x HLVA_PROMI
Align seg 1/1 to: HLVA_PROMI from: 1 to: 1577

40 AAAGCCCTAAACCGCGCGCATCGCTTCGCGCGCTTACTTACGCAAT 89
||| ..... ||| ||| ||| |||
7 LysLeuSerProSerGlyArgLeuAlaIleSerLeuAlaIleLeuPheVa 23
ATGCTGTGCTTC.....GGCATTTGCGCCAGCGCGCGGAGAC 130
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
23 LserLeuAsnAlaIleTYRGLYAsnGlyIleValProAsp.....AlaGly 38
131 AC..... 132
|||
38 IseGlnGlyProAspValSerAlaValAsnGlyGlyThrGlnValIleAsn 54
133 .....ACTTATTCGGCATCACTACCAATACATGCGCACTT 170
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 IleValThrProAsnAsnGlnGlyIleSerHISAsnGlnTYRTrGlnAsp 71
171 TGCCGAAATTAAGGCAAG...TTTGCACTGGGCGGCAAGATATTTGAG 217
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 e.....AsnValGlySerProGlyAlaValPheAsnAsnAlaLeuGln 86
218 TT.....TACACAAAAAGGGAGTTGCTC 243
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 IacGlyInSerGlnLeuAlaGlyHISLeuAsnAlaAsnSerAsnLeuAsn 102
244 GCGAAATCGATGACGAAGCCCGATGATGATTTTGTGTGATTCGGG 293
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
103 GlyGlnAla.....AlaSerLeuIleLeuAsnGlnValAlaSerAr 116
294 TAAC.....GGCGTGGCGGCAAT 310
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 GAsnProSerPheLeuLeuGlyGlnGlnValAlaPheGlyIleAlaAla. 132
311 TGCGGGCGCATCATATATTGTGAGCGTGCGACATAC..... 348
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 .....GlyTYRValLeuSerAsnProAsnIleIleThrCysAsp 145
349 .....GGCGGCTATAACAATGTTGATTTGGTGGAGAGCAAGCAATCC 392
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 GlyCysGlyPheThrLeuAsnThrSerArgSerSerLeuValGlyAsnTr 162
393 C.....GATCACACCGCGCTTTCTTACCAAAATTGGAAGAATA 433
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
162 OleuPheGlnAsnGlyGlnLeuLysGlyTYRTrHISLeuAsnThrAla 179
434 AT...TATAAGCAGGACTAACGGCATCTTATGCGCGG.....GAT 474
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 snLeuLeuSerLeuGlyLysAsnGlyLeuAsnThrThrGlyLeuLeuAsp 195
475 TATCATATATCCGCGGTTTGCAAAA.....TTTGACAGATGACGAAC 518

```



```

1831 .AATTACCAACCGGAGAACGGGATCGACTTTACTGCTTTCCGGCGGAA 1879
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
762 nLysLysGlnThrAspAsnThrSerThr.....IleSerGlyGlyP 777
1880 CAATTAAACGGCAATATCACCAACAACGCGCAACTGTTTTCACG 1929
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
777 heSerThrGlyGlyValAspLysValGlySerLysAlaAspPheGln 793
1930 GCGAGACCGACCGCGCTACATCATTTAGGAGCGGGTGGTCAAA 1979
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
794 .....TyrAspLysGlnHisThrGlnThrGluValThrLys 805
1980 AATGAGAGTATCCACAGAGAAATCGTGGGAC..... 2016
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
805 sAsnAlaGly...SerGlnThrGluValAlaGlyAspLeuThrIleThrAla 821
2017 .AACGATTGGATCGACCGCACATTAAAGCGGAAATTCATATTCAG 2064
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
821 lAsnLys.....AspLeuLeuHisGlyGlyAlaSerHisHisValGlu 835
2065 GCGGACAGCGGTGGTTCGCCGCAATGTGCCAAAGTGGAGGGGATGG 2114
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
836 GlyArgTyrGlnGlnSerGlyGlnLysIleGln..... 846
2115 GCATTTAAGCAATCACGCCCAAGCATTTTCGCTGCGCACCGCATCAAA 2164
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
847 .HisLeu.....AlaValAsnAspS 853
2165 GCGACACATCTGTACACGTTG..... 2187
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
853 ergLThrSerLysThrAspSerLeuAsnValGlyIleAspValGlyVal 869
2188 .....GACTGACGGGTCTGCAAGTTGTACCGAAACCAATCCAGCA 2231
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 AsnLeuAspTyrSerGlyValThrLysProValLysLysAlaIleGlnAs 886
2232 CGAT...AAGTATTGCTTCATTAGACAGACGACATCAGAGGCAATG 2278
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
886 pGlyValAsnThrThrLysProGlyAsnAsnThrAspLeuThrLysLysV 903
2279 TCAGCCTTCCGATCAC...GCTCATTTA...AATCTACAGGATTTGCC 2322
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
903 alThrAlaArgAspAlaIleAlaAsnLeuAlaAsnLeuSerAsnLeuGlu 919
2323 ACACTCACGCGCAATCTTAGTCAGCG..... 2349
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
920 ThrProAsnValGlyValGluValGlyIleLysGlyGlySerGlnG 936
2349 ..... 2349
936 nSerGlnThrAspSerGlnAlaValSerThrSerIleAsnAlaGlyLysI 953
2350 .....GAGACACGACATATACG 2367
953 leAspIleAspSerAsnAsnLysLeuHisAspGlnGlyThrHisThr... 968
2368 GTTACGCGCAACGCCCAACGCGCAACTCAGCTCTGGGCAATGC 2417
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
968 .....GlnSerThrGlnGlnGly...IleSerLeuThrAlaAsnThr 981
2418 C.....CAAGCAATTTAATCAG 2437
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
981 HisThrSerGlnAlaThrLeuAspLysHisGlnThrThrPheHisGlnT 998
2438 CCACATTAAACGCAACACATCGCTCGGACAAATGCTTCATTAAATCA 2487
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
998 hLysGlyGlyGlnIleGlyValSerThrLysThrGlySerAspIle 1014
2488 AGCAACAACGCGGTACAAACGCGCATGCTGACGTTTCCGCAACGCT. 2535
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1015 ThrValAlaIleLysGlyGlnGlyGlnThr.....AspAsnAlaLe 1029
2536 .....AAGCAAAAGTAAGCAATTCGCGCACTACAGCGCAATGCTCT 2575
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1029 uMetLThrLysAlaLysGlySerGlnPheThrSerAsnLysAspLies 1046
2576 CCCTAGCCGATTAAGCAGTATTCATTGTGAAAACAGCCGCTT. 2619
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1046 erLleAsnValGlyGluAsnAlaHisThrGlyGlnGlyAlaGlnPheAspAla 1062
2620 ...ACCGGAAA.....ATCAGCGCGCGCAAGGATACGGCATTAACATT 2660
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1063 GlnLysGlyLysThrValIleAsnAlaGly..... 1072
2661 AAAGACACGCAATGACGCTGCCGTGGGACGAGATTAAGCAATTTAA 2710
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1073 .....GlyAspLeuT 1076
2711 ACCTTGACACCGCCACCATTACTCAATTCGCTATCAGACAGATGCG 2760
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1076 hLysAlaGlnAlaThr..... 1081
2761 GCAGGCGCGCAACCGGACGATGCGGACAGATGCGCGCGCGTTCGCG 2810
1081 ..... 1081
2811 CGGTTCCCTATTATCCGTTACGCGCGCACTGGCGAGAAATCCGTTCA 2860
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1082 .....AspThrHisSerGlnSerGlnSerAla 1090
2861 ACACGCTGACGTAACCGCAATTTGAACGGTCAAGGACATTCGCTTT 2910
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1090 sn.....ValAsnGlySerAlaAsn..... 1096
2911 ATGTGCGACACTTCGCGCTACCGGACGCGCAATTTGAAGCTGGCGAAAG 2960
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1097 .....LeuLysValGlyLysThr 1102
2961 TTCGAGGACCATTTACACCTTGCGCTGCACAAATATCCGAGCAAGACCG 3010
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1102 rProGlnSerLysAspTyrGlyGlyGlyPheAsnAlaGlyThrThrHis 1119
3011 TAACTTCGCAATTTAGC.....GTAGTGAAGGAAAAAGAC 3048
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1119 lSerLysGlnGlnThrThrAlaLysValGlyThrIleThrGlySerGln 1135
3049 AACACACCGCTGTCGCAAAATCTTAATTCACCCCTGCAAAACACACGT 3098
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1136 GlyIleGlnLeuAsnAlaGlyHisAsnLeuThrLeuGlnGlyThrHisLe 1152
3099 CGATGCCGCGCATGGCGTTATGAGTTATCCGCAAAAGACGGAGTTC 3148
1152 u..... 1152
3149 GCTTGCAATAATCCGGTCAAGAACAGAGCTTTCGACAAACTCGGCAAG 3198
1152 ..... 1152
3199 GCGGAGAAACAGAGCGCGCTTGACGCGCAAA.....CAGGACAC 3239
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1153 SerSerLysGlnAspIleAlaLeuAsnAlaThrAsnLysValAspLeuG 1169
3240 ACTTGCGCGCAACACAGCGGAGGAAAAAGACAGCGCAAGGCTTGACG 3289
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1169 nSerAlaSerSerGlnHisThrGlnLysGlyAsnAsnLeuSerGlyGly 1186
3290 CGCTGATTGGGCGCGGCGCAATGCGCAAGAAAAAGCAGAAATGTT... 3336
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1186 alGlnAlaGlyPheGlyLysLysMetThrAspAlaSerValAsn 1202
3337 .....GCGGACCGGCG 3347
1203 GlyLeuGlySerAlaGlnPheAlaIleGlyLysGlnAspGlnLysSerVal 1219
3348 CCGGACGCGAGCGCGG.....GAAAATGCCGCGCATTTATGACGCGGAG 3391

```





```
1366 AACGACCGCCCTGTCCAAATCGCAAGGC.....ACGCT 1400
|||||
2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThr 18
1401 GCTGTTTCAGCAAGCCAAAGG..... 1419
|||||
18 rLeuAlaMetValLeuGlyAlaLeuGlyAlaAlaProAlaAlaTyrAla 35
1420 .....GAAGACCAAGGCTCGGTCAAGCTGGCGGCAAGCTAA...GTCATC 1461
|||||
35 sPrtPasnAsnGlnSerIleIleLysAlaGlyLeuArgGlnHisGlyIle 51
1462 TTACATCAGCAGCGGAGCATCAAGCAAAAAACACGCTTTGATGAAAT 1511
|||||
52 HisIleLysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThr 68
1512 CGGTTGTCAGCGGAGGCGGAGCGGTGCACTGATCGCATATTCAGT 1561
|||||
68 rIleLysValSerGlyArgGlnAlaGlnGlyValLeuLeuGlnAsn... 83
1562 TCAACCCCGCAAACTCTATTTCGGCTTCGGCGGAGAGCTTGGATTC 1611
|||||
84 .....ProAlaAlaGlnLeuArgPheGlnAsnGlySerValThrSer 97
1612 AACGGCATTTCGCTTCGTTCCACCGCATTCAAATATCCGATGAGGCG 1661
|||||
98 SerGly.....GlnLeuPheAspGlnGlyVal 106
1662 GATGATTGTCAACACATCAAGCAAGCAAGATCCACCGTTCACATTACG 1711
|||||
106 I.....ArgArgPheLeuGlyThrValThrValLys 117
1712 GCATATAA.....GATATTACTACACCGGCAATTAACAACACTTG 1752
|||||
117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaSerAspThr 133
1753 GATACCAAAAAAGAAATTGCC...TACAAACGGTTGGTGGCGAGAAAGA 1799
|||||
134 ArgAspAspAspGlyIleAlaLeuTyrValAla.....GlyGlnGln 147
1800 TGCACCAAAACGACGCGGCGGCTCAATCTGAATTACACCGGAAAG 1849
|||||
148 .....AlaGlnAlaSerIle 153
1850 CGGATCGCACTTACTGCTTCGCGGGAACAATTTA.....AACGCG 1893
|||||
153 laAspSerThrLeuGlnGlyAlaGlyValArgValGlnArgGlyAla 169
1894 AATATGACCAAAACAAACGCAACTGTTTTCAGCGGCGACACGACACC 1943
|||||
170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu..... 183
1944 GCACGCGCTCAATCATTTAGAAAGCGGCGTGC...AAATGGAAGTA 1990
|||||
184 .....HisIleGlyThrLeuGlnProLeuGlnProGlnAspL 196
1991 TCCCAACAGAGAAATCGTGGGACAACGATTTGATGACCGC..... 2034
|||||
196 euProProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212
2035 .....ACATTTAAAGCGGAAACTTCCA 2057
|||||
213 AlaSerGlyAlaProAlaAlaValAlaPheValPheGlyAlaAsnGlnLeuThr 229
2058 TATTCAGGCGGACAGCGGTGGTTTCCCGCAATGTGCAAAAGTGAAG 2107
|||||
229 rValAspGlyGly..... 233
2108 CGGATTGCGATTAAACAATCAGCCCAAGCAGATTTCGGTGGCGACCG 2157
|||||
234 .....HisIleThrGlyGlyArgAlaAla.....GlyValAla... 244
```

```
2158 CATCAAAAGCCACACAATCTGTACAGCTTCGAGTGGAGCGGTGTACAAG 2207
244 ..... 244
2208 TTGTACCGAAAAAACCATTAACGAGATTAAGTATGTTGTTTATGAGCA 2257
|||||
245 .....AlaMetAspGlyAlaIleValHisLeuGln 255
2258 AACCCGACATCAGAGGCAATGTACAGCTTCGCCATCAGCTCATTTAAT 2307
|||||
255 rGlaThrIleArg..... 259
2308 CTCACAGAGCTTCCACACTCAGCGCAATCTTACTGACGCGGAGAC 2357
|||||
260 .....ArgGlyAspAlaProAlaGlyGly..... 267
2358 GCACATACGCTTACGCGCAACGCCCAAAACGCAACCTC..... 2400
|||||
268 .....AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282
2401 .....AGC 2403
282 lYPheGlyProLeuLeuAspGlyTrpTyrGlyValAlaSerAspSer 298
2404 CTCGTGGGCAATGCCCAAGCAACATTTATCAAGCCACATTTAAACGCAA 2453
|||||
299 ThrValAspLeuAlaGlnSerIleValGlnAlaProGlnLeuGlyAla 315
2454 CACATCGGCTTCGACAAATGCTTCATTTAATCTAAGC..... 2490
|||||
315 alIeArgAlaGlyArgGlyAlaArgValThrValSerGlyGlySerLeu 332
2491 .....AACACCGCGTACAAAAGCGC..... 2511
|||||
332 erAlaProHisGlyAsnValIleGlnThrGlyGlyAlaArgArgPhe 348
2512 .....AGTGTACGCTTCGCAACGCGCTAAGGC 2540
|||||
349 ProProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArg 365
2541 AAAC.....GTAAAGCATT 2554
|||||
365 agIleGlyArgAlaLeuLeuTyrArgValLeuProGlnProValLysLeu 382
2555 CCGCACTCAAGGCAATGTCCTCCATAGCCGATAGCAAGTATTCATTT 2604
|||||
382 hrLeuAlaGlyGlyAlaGlnGlyGlnGlyAsp.....IleValAlaThr 396
2605 GAAACAACGCGCTTACCGGAAATAATCAGCGCGGCAAGATTAAGCAT 2654
|||||
397 GlnLeuProProIleProGlyAlaSerSerGlyProLeuAspValAla 413
2655 ACACTTAAAGACAGCGAATGACGCTGCGGCGGACGGAATTAGCA 2704
|||||
413 u...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAlaValAsp 428
2705 ATTTAAACCTTGACACACCAATTTACACTCATTTCCGCTATCGCAC 2754
|||||
428 erLeuSerIleAspAspAlaThrTrpValMetThrAsp..... 440
2755 GATCGGCGAGCGCGCAACCGGACGATGCGGCGGCGCGCGCG 2804
|||||
441 .....AsnSerAsnValGlyAlaLeuArgLeuAl 450
2805 TTGCGCGCGTTCCTATATTCGTTACGCGGCACTTCGCGCAATCC 2854
|||||
450 aserAspGlySer...ValAspPheGlnGlnProAlaGlnAla...Gly 465
2855 GTTTCACACGCTGACGTTAAAGCAAAATTTGAAGTCAAGTCAACATTC 2904
|||||
465 rPheLysValLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPhe 480
2905 CGCTTATGTGGAACCTTCGGCTACCGGACGCGCAAAATTTGAAGCTGCG 2954
```



OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-44.  
 RC STRAIN-ISOLATE CN7531;  
 RA MEDLINE=92407514; PubMed=1527510;  
 RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.,  
 RT "Cloning, nucleotide sequence and heterologous expression of the  
 RT protective outer-membrane protein P.68 pertactin from Bordetella  
 RT bronchiseptica";  
 RL J. Gen. Microbiol. 138:1697-1705(1992).  
 CC -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS  
 CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN  
 CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.  
 CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)  
 CC CONCENTRATIONS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X54815; CAA38584.1; -;  
 DR EMBL: A19180; CAA01453.1; -;  
 DR PIR: A47675; A47675.  
 DR InterPro: IPR003992; pertactin.  
 DR InterPro: IPR003991; pertactin\_related.  
 DR PRINTS: PR01482; PERTACTIN.  
 DR PRINTS: PR01484; PERTACTINFAMILY.  
 KW Outer membrane; Signal; Virulence; Repeat.  
 FT SIGNAL 1 34  
 FT CHAIN 35 911 P.94.  
 FT CHAIN 35 712 PERTACTIN (P.68).  
 FT PROPEP 713 911 POTENTIAL.  
 FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN  
 CELL ATTACHMENT TO VARIOUS EUKARYOTIC CELL  
 LINES).  
 FT SITE 701 703 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DOMAIN 266 275 3 X 5 AA TANDEM REPEATS OF G-G-A-V-P.  
 FT REPEAT 266 270 1.  
 FT REPEAT 271 275 2.  
 FT REPEAT 276 280 3 (APPROXIMATE).  
 FT REPEAT 280 280 7 X 3 AA REPEATS OF P-Q-P.  
 FT DOMAIN 570 601  
 SO SEQUENCE 911 AA; 93995 MW; 307BDF6EC2D987A1 CRC64;

alignment\_scores:  
 Quality: 274.00 Length: 1149  
 Ratio: 0.584 Gaps: 50  
 Percent Similarity: 40.818 Percent Identity: 20.540

alignment\_block:  
 US-09-303-518d-653 x PERT\_BORBR ..

Align seg 1/1 to: PERT\_BORBR from: 1 to: 911

1306 GGCGGGGGCTCATATCATGATGAGTACCGTTACTTGGAAAGTAAA 1355  
 |||||  
 58 G1YAlaG1YValArgThrAlaThrIle.....LysValSe 72  
 1356 CGGCGTGGCAAGACCGCTGTCCAAATCGCAAGACGAGCGTGTGG 1405  
 |||||  
 72 rGly.....ArgIAlaG1ng1YValLeuLeug 82  
 1406 TTCACGCCAAAGGGAACCAAGCGTGGTACGCTGGCGGCGGTAAA 1455  
 |||||

82 LuAsnProAlaIaGlu.....LeuArgpHeGInAsnGlySer 94  
 1456 GTCAATCTTAGTACGACGCGGAGATCAAGCGCAAAACAAACGCTTAG 1505  
 |||||  
 95 ValThrSerSerGlyGlnLeuPheAspGlnGlyValArgpPheLeu 110  
 1506 TGAATCGGCTTGGTCAAGCGGAGGAGGAGGTCGCAATCAATCCGATA 1555  
 |||||  
 111 .....GlyThrValThrValLysAla.... 117  
 1556 ATCAGTTCACCCCGCAAACTGATTTCGGCTTTCGGCGGACGTTTG 1605  
 |||||  
 118 .....GlyLysLeu 120  
 1606 GATTGAAGGCGCATTCGCTTTCGTTCCACCGCATCAAAATCCGATGA 1655  
 |||||  
 121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgspAspAs 137  
 1656 AGGGCGGATGATGTCAACCAACCAATCAAGACMAAATCCACGCTT... 1701  
 |||||  
 137 pGlyIleLeuLeuThrValAlaGlyGlnAlaGlnAlaSerIleAla 154  
 1702 .....ACCATTTACGCAATTAAGATATTTACTACACCGCATATACAC 1746  
 |||||  
 154 spSerThrLeuGlnGlyAlaGlyValArgValArgValArgValArg 170  
 1747 AACTTGATACCAAAAGAAATGCTTACACGCTTGTGGCGCGGAA 1796  
 |||||  
 171 ValThrValGlnArg..... 175  
 1797 AGATGCAACCAAAAGACGCGGCTCAATCTGAAT.....T 1834  
 |||||  
 176 .....SerThrIleValAspGlyGlyLeuHisIleGlyThrLeuGlnPro 191  
 1835 ACCAACCGGGAAGAAGCGGATCGCACTTTACTGCTTCGCGGCAACAAT 1884  
 |||||  
 191 eugInProGlnAspLeuProSerArgValValLeuGlyAspThrSer 207  
 1885 TTAAACGGCAATATCAACCAACCAAGCAACTGTTTTCAGCGGACG 1934  
 |||||  
 208 ValThrAla.....ValProAlaSerGly..... 215  
 1935 ACCGACACCGCAGCGCTTACATTTAGGAGAGCGGTGTCAAAATGG 1984  
 |||||  
 216 .....AlaProAlaIleValSerValPheGlyAlaAsnGlnLeuThrVal 231  
 1985 AAGTATTCACCAAGAGAAATCGTGGGACACGATTCGATCGACCGC 2034  
 |||||  
 231 spGly..... 232  
 2035 ACATTTAAGCGGAAACTTCATATTCAGGGGCGCAACGCGTGTTC 2084  
 |||||  
 233 .....GlyHisIleThrGlyValArgAlaAla.... 241  
 2085 CCGCAATGTTGCCAAAGTGAAGCGATGGCATTTAAGCAATACGCGCC 2134  
 |||||  
 242 .....GlyValAlaIleAlaMetAspGly..... 248  
 2135 AAGCATTTTCGGTGTGCAACCGCATCAAGACACACATCTGTACAGT 2184  
 |||||  
 248 ..... 248  
 2185 TCGAGCTGAGCGGCTGACAACTTTAGCGAAACCAATACCGACGA 2234  
 |||||  
 248 ..... 248  
 2235 TAAAGTATGCTTCATTTAGCAAGACGACATCAGAGCAATGTGAGCC 2284  
 |||||  
 249 .....AlaIleValHisLeuGlnArgAlaThrIleArg..... 259  
 2285 TTGCCGATCAACGCTCATTTAAATCTCAGAGGACTTGGCACACTCAACGGC 2334  
 |||||  
 260 .....ArgGly 261



```

719 ValGlyValYtrAlaThrYtrIleAlaHisSerGlyPheThyIleuAspAl 735
3940 .....TCAG 3943
735 arthleuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySerA 752
3944 ACGGC.....ATCAGAGCGCAAAATCCGCCGCCCGCGTGCATTCACGC 3987
||||| .....||||| ||| .....|||||
752 spGlyYtrAlaValLysGlyLysYtrArghr.....HisGly 764
3988 ATTCAGCAAGATATCCGCGAGGT.....TCGCGCGATT 4022
::: ||| .....||||| ||| .....|||||
765 ValGlyAlaSerLeuGluAlaGlyArgArgPheAlaHisAlaSpGlyTr 781
4023 CGGCATTCGACCGCACATCCGGCGACGCGCTATTTCGCCAAAAAGCG 4072
: .....|||||: : : : : ||| : : : :
781 pPheLeuGluProGlnAlaGluLeuAlaValAlaPheArgValGlyGlys 798
4073 ATTCAGCATACGCAAAAC..GTCAATATCCGCACCCCGGCCCTTCGATTTC 4119
:::||||| ||| .....: : : : : ||| : : : :
798 eYtrYtrGAlaAlaHisnGlyLeuArghValAlaArgAspGluGlySerSer 814
4120 AACCGCTACCGCGCGCGCATTAAGACGATTATTCATTCGCAACCGCGCA 4169
||| .....|||||: : : : : ||| : : : :
815 ValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaG 831
4170 ACACATTCATCATCAGCCTTATTAGCCTGCTCTATACGATGCCGCTT 4219
: : : : : ||| .....: : : : : ||| : : : :
831 yGlyArgGlnValGlnProGlyTrIleLysAlaSerValLeuGlnGluPheA 848
4220 CGGCGCAAGTCCGACGCGCGCTATACGCCGCTATTGGCGCAGATTTC 4269
: : : : : .....|||||: : : : : ||| : : : :
848 spGlyAlaGlyTrValArghTrHisnGlyLeAlaHisArghTrGluLeu 864
4270 GGCAAACCCGCGATCGCGAATGGCGGTAAACGCCAAATCAAGATT 4319
||||| ||||| ||||| |||
865 ArgGlyThrArg...AlaGluLeuGly..... 872
4320 CACGCTGTCCTCCACGCTGCGCGCCGACGAGGGCGCAATTGAGACG 4368
|||||: : : : : |||||: : : : : |||
873 ...LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuYtrAlaAs 888
4369 .....CAGCACAGCGCGGCATCAATTAGCAGCGCGG 4404
: : : : : ||| ||| .....: : : : : |||
888 eYtrGluYtrSerLysGlyProLysLeuAlaMetProTrp 901
seq_name: SwissProt_40:PERR_BORPE
seq_documentation_block:
ID PERT_BORPE STANDARD: PRT: 910 AA.
AC p14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Peptactin precursor (Outer membrane protein P.69) (P.93).
GN P69 OR OMP69A.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RA [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN-ISOLATE CN292;
RX MEDLINE=89264462; PubMed=2542937;
Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
Novotny P., Morrissey P., Fairweather N.F.;
"Novotny P. Cloning and characterization of protective outer
protein P.69 from Bordetella pertussis.";
RT Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
RL [2]
RP REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;

```

```

RA  Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT  "Cloning nucleotide sequence and heterologous expression of the
RT  protective outer-membrane protein P.68 pertactin from Bordetella
RT  bronchiseptica."
RL  J. Gen. Microbiol. 138:1697-1705(1992).
RM  [3]
RP  X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RA  MEDLINE=96196517; PubMed=860998;
RX  Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT  "Structure of Bordetella pertussis virulence factor P.69 pertactin.";
RL  Nature 381:90-92(1996).
CC  -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC  MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC  BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC  -1- SUBUNIT: MONOMER.
CC  -1- SUBCELLULAR LOCATION: Outer membrane.
CC  -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC  -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC  CONCENTRATION.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch)
CC  -----
DR  EMBL: J04560; AAA22980.1; ALT_SEQ.
DR  PIR: A32560; A32560.
DR  InterPro: IPR003992; pertactin.
DR  InterPro: IPR003991; pertactin_related.
DR  PRINTS: PRO1482; PERTACTIN.
DR  PRINTS: PRO1484; PERTACTINFAMILY.
KW  Outer membrane; Signal; Virulence; Repeat.
FT  SIGNAL 1 34
FT  CHAIN 35 910 P.93.
FT  CHAIN 35 711 PERTACTIN (P.69).
FT  PROPEP 712 910 POTENTIAL.
FT  SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
FT  ADHESION TO VARIOUS EUKARYOTIC CELL
FT  LINES).
FT  DOMAIN 266 290 5 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT  REPEAT 266 270 1.
FT  REPEAT 271 275 2.
FT  REPEAT 276 280 3.
FT  REPEAT 281 285 4 (APPROXIMATE).
FT  REPEAT 286 290 5 (APPROXIMATE).
FT  DOMAIN 579 593 5 X 3 AA TANDEM REPEATS OF P-Q-P.
SQ  SEQUENCE 910 AA; 93452 MW; A169671E20A2E7DB CRC64;

Alignment_scores:
Quality: 270.00 Length: 1156
Ratio: 0.554 Gaps: 53
Percent Similarity: 42.128 Percent Identity: 21.367

Alignment_block:
US-09-303-51BD-653 x PERT_BORPE ..

Align seg 1/1 to: PERT_BORPE from: 1 to: 910

1366 AACGACGCCCGTGTCCAAATCGGCAAGGC.....ACGCT 1400
||||| ||||||| |||
2 ASmetSerLeuSerAlaArgIleValIysAlaIaProLeuAlaArgAlaThrIth 18
1401 CGTGGTTCAGCCAAAGG..... 1419
||||| |||
18 rIeuAlaMetAlaLeuGlyAlaLeuGlyAlaIaIaProAlaIaIaHisAlaA 35
1420 .....GAAGACCAAGCGTCGGTCAGACCGTGGCGAGCGTAAGATCACTTA 1464
||||| ||||||| |||
35 sPTPrAsnAlaGlnSerIleValIysThrIleGlyAlaIaGlnIaHisGlyIle 51

```

```

1465 GATCAGCAGCGCGGATCATCAAGCAAAAAAACACCTTTAGTGAATCGG 1514
      |||:::|||||
52 HistleuInglySerAspProglyValArgThrAlaSerGlyThrTh 68
      |||:::|||||
1515 CTGG...GTACGGGCGAGGGGCGGTGCAATGCAATGCCGATATATGAGT 1561
      |||:::|||||
68 rIleuValSerGlyArgGlnAlaGlnGlyIleuLeuGlnAsn... 83
      |||:::|||||
1562 TCAACCCCGCAACCTCTATTTCGGCTTTCGGCGGAGCGTTGGATTTCG 1611
      |||:::|||||
84 .....ProAlaAlaGlnLeuGlnPheArgAsnGlySerValThrSer 97
      |||:::|||||
1612 AACGGGCAATTCGCTTTCGCTTCACCGCATTCMAATATCCGATGAAGGCG 1661
      |||:::|||||
98 SerGly.....GlnLeuSerAspAspGlyIle 106
      |||:::|||||
1662 GATGATTGTCAACCAATCAAGCAAAAAAGAAATCCACGCTTACCTATCAG 1711
      |||:::|||||
106 e.....ArgArgPheLeuGlyThrValThrValLysa 117
      |||:::|||||
1712 GCATTTAA.....GATATTACTACACCGGCAATTAACACACTTCG 1752
      |||:::|||||
117 IagIlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValGlyAspThr 133
      |||:::|||||
1753 GATGCAAAAAAGAAATTCCTACACGCGTTGGTTGGCGGAAAGATGC 1802
      |||:::|||||
134 .....TrrAspAspAspGlyIleAl 140
      |||:::|||||
1803 AACCAAAACGACGCGCGCTCATCTGMAATTACCAACCGGAGACGCG 1852
      |||:::|||||
140 AleuTyValAlaGlyGln.....GlnAlaGlnAlaSerIleAla 154
      |||:::|||||
1853 ATCGCACTTACTGCTTCGCGGCGGACAAATTA.....AACGGCAAT 1896
      |||:::|||||
134 spSerThrLeuGlnGlyAlaGlyValGlnIleGluArgIlyAlaAsn 170
      |||:::|||||
1897 ATCAGCAAAACGCAAACTGTTTTCAGCGGACGACGACGACGCA 1946
      |||:::|||||
171 ValThrValGlnArgSerAlaIleValAspIlyLeu..... 183
      |||:::|||||
1947 CGCCTCAATCATTTAGAGAGCGGCGTGTCA...AAAATGAGAGTATCC 1993
      |||:::|||||
184 .....HisIleGlyAlaLeuGlnSerLeuGlnProGlnAspLeuP 197
      |||:::|||||
1994 CACAAGAGAAATCTCTGGGACAAAGATGGATGACCGGACA..... 2037
      |||:::|||||
197 rOfProSerArgValAlaLeuArgPThrAsnValThrAlaValProAla 213
      |||:::|||||
2038 .....TTTAAAGC 2045
      |||:::|||||
214 SerGlyAlaProAlaAlaValSerValLeuGlyAlaSerGlnLeuThr 230
      |||:::|||||
2046 GGAAGAACTTCATATTTCAGGCGGACAAAGCGGTGTTCCCGCAATGTTG 2095
      |||:::|||||
230 uAspIlyGlnHisIleThrGlnGlyArgAlaAla.....GlyVala 244
      |||:::|||||
2096 CCAAGTGTGAAGCGGATGGCATTTAACAATCAAGCCCAAGCAGTTTTC 2145
      |||:::|||||
244 Ia.....AlaMetGln 247
      |||:::|||||
2146 GGTGTGCAACCGCATCAAGCCACAAATCTGTACACCTTCGAGC..... 2190
      |||:::|||||
248 GlyAlaValAlaHisLeuGlnArgAlaThrIleArgArgGlyAspAlaPr 264
      |||:::|||||
2190 ..... 2190
      |||:::|||||
264 oAlaGlyAlaValAlaProGlyGlyAlaValAlaProGlyAlaValAlaProG 281
      |||:::|||||
2191 .....TGACGGGT 2199
      |||:::|||||
281 IyGlyPheGlyProGlyGlyPheGlyProValLeuAspGlyTrrPlyrGly 297
      |||:::|||||
2200 CTGACAAAGTTGTACCGAAAAAACCATTAACGACATTAAGTATGCTTC 2249
      |||:::|||||
298 Val..... 298
      |||:::|||||
2250 ATTGAGCAAGACCGACATCAGAGC...AAGTCAAGCTTCCGATCAG 2296
      |||:::|||||
299 .....AspValSerGlySerSerValGlnLeuAlaGln... 309
      |||:::|||||
2297 CTCATTTAAATCTCAAGGACTTGGCCACATCAAGCGCAATCTAGTCA 2346
      |||:::|||||
310 .....SerIleValGlnAlaProGlnLeuGlyAlaAlaIleArgVal 323
      |||:::|||||
2347 GCGGAGACACCGCACTATCGGTTACGGCAAC.....GCCACCCAAA 2390
      |||:::|||||
324 GlyArgGlyAlaArgValThrValSerGlyLysSerLeuSerAlaProH 340
      |||:::|||||
2391 CCGCAACTCTACGCTCGTGGCGAATGCCAAGCAATTTAATCAAGC... 2439
      |||:::|||||
340 sGlyAsnValIleGluThrGlyAlaArgArgPheAlaProGlnAla 357
      |||:::|||||
2440 .....ACATTAACGGCAACATCGGCTTCGACAAATGCT 2475
      |||:::|||||
357 IaProLeuSerIleThrLeuGlnAlaGlyAlaHisAlaGlnIlyAla 373
      |||:::|||||
2476 TCATTTAATCTAAGCAACACCGCGTACAAAGCGAGTCTGACGCTTC 2525
      |||:::|||||
374 Leu.....LeuTyArgValLeuProGlyProValLysLeuThrLeuTh 388
      |||:::|||||
2526 CGACAAAGCTAAGCAAAAGTAAAGCATTCGCCACATCAAGCGCATGCT 2575
      |||:::|||||
388 rGlyGlyAlaAspAla.....GlnGlyAspIleVal 398
      |||:::|||||
2576 CCTTACCGCATTAAGCGCATTTCCATTTTGAACACGCGCTTACCGGA 2625
      |||:::|||||
398 AlaAlaThr.....GlnLeuProSerIleProGly 407
      |||:::|||||
2626 AAATATACGCGCGGCAAGATACGATACACTTAAAGACACGGAAG 2675
      |||:::|||||
408 ThrSerIleGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTr 423
      |||:::|||||
2676 GACGCTGCGGTGGGCAAGCAATTAAGCAATTTAACTTGACACGCA 2725
      |||:::|||||
423 PThr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaTr 439
      |||:::|||||
2726 CCATTACACTCAATTCGCTATTCAGACAGATGCGGACGCGGCAACC 2775
      |||:::|||||
439 hTrPValMetThrAsp..... 444
      |||:::|||||
2776 GGCAGTGGCGGACAGATGCGCGCGCGCGCTTCGCGCTTCCCTATTATC 2825
      |||:::|||||
445 AsnSerAsnValGlyAlaLeuArgAlaSerAspLysr...ValAs 460
      |||:::|||||
2826 CGTTACGCGCGCAACTTCGCGACAAATCCGTTTCAACACGCTGACGGTAA 2875
      |||:::|||||
460 PheGlnGlnProAlaGlnAla...GlyArgPheLysValLeuThrVala 476
      |||:::|||||
2876 ACGGCAAAATGAACGGTCAAGGAAATTCGCTTATGTGCGAACTCTTC 2925
      |||:::|||||
476 sn...ThrLeuAlaGlySerGlyLeuPheArgMetAsnValPheAlaAsp 491
      |||:::|||||
2926 GCGTACCGGCGGCAATTTGAAGCTGCGGGAAGTTCCGAAGCAGCTTA 2975
      |||:::|||||
492 LeuGlyLeuSerAspLysLeuValValMetGlnAspAlaSerIlyGlnH 508
      |||:::|||||
2976 CACCTTGGTGTCAACATACCGGCAACGAAACCGTAAGTCTTCGACCAAT 3025
      |||:::|||||
508 sArgLeuTrpValArgAsnSerGlySerGlnProAlaSerAlaAsnThrL 525
      |||:::|||||
3026 TGACGGTAGTGAAGAAAAAGACAACACACCGCTGTCCGAAATCTTAAT 3075
      |||:::|||||
525 euLeuLeuValGln.....ThrProLeuGlySerAlaAlaThr 537
      |||:::|||||
3076 TTCACCTGCAAAAAAGAA.....CAAGTGAATGCGGCGCAATGGGCTTA 3119
      |||:::|||||

```

```

|||||  |||||  |||||  |||||  |||||
538 PheThrLeuAlaAsnLysAspGlyLysValAspIleGlyThrTyrArgTyr 554
3120 TCAGCTTATCCGCAAGAGCGGCGATTCGCCCTGCATTAATCCGGTCAAG 3169
554 TrgLeu..... 556
3170 AACAGAGCTTTCCGCAAACTCGCAAGGCGGAGAAACAGAGCGCGCC 3219
556 ..... 556
3220 TTGACGGCAAAACAGCACAACCTTCCGCCAACAACAGCGGAAAGA 3269
557 ..... 557
3270 CAACGGCAAAAGCCTTGACGGCGTGTATGGCGCGCGCGGCATGCCACCG 3319
|||||  |||||  |||||  |||||  |||||
560 yasnGlyIntP.....SerLeuValGlyAla..... 569
3320 AAAAGCAGAAAGTGTGGCGAAGCGCGCGCGAGCGGCGGAAAT 3369
|||||  |||||  |||||  |||||  |||||
570 ..LysAlaProProAlaProLysProAlaProGlnProGlnProGln... 584
3370 GCGGCGATTATGACGGCGAGAGAGAGAAAGAGGGTGCAGCGCGATTA 3419
584 ..... 584
3420 AGACACCGCTTGGCAAAACAGCGGAAAGCGAAACCGCGCGGCTACCA 3469
585 ..... 585
3470 CCGCCTTCCCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGAG 3519
589 ..... 589
3520 CCCCA.....CCGCAACCCCAACCGCGCGCGCGCGCGCGCTATCAGCG 3560
|||||  |||||  |||||  |||||  |||||
593 ProGlnAlaProAlaProGlnProProAlaGlyArgGlyLeu...SerAl 608
3561 TTATGCCAATAGCGGTTGAGTGAATTTTCCGCGCAAGCTCAACAGCTT 3609
|||||  |||||  |||||  |||||  |||||
608 aAlaAlaAsnAlaAlaValAlaAsnThrGlyGlyAlaGlyLeuAlaSerThr 625
3610 ....TTGGCGCTACAGCAGCAATTTGCGCGCGGCTTTGCCGAGACCGC 3654
|||||  |||||  |||||  |||||  |||||
625 eutTrpTyrAlaGlnSerAlaValSerLysArgLeuGlyGlyLeuArg 641
3655 CGCAAC.....GCCGTTGGCAAGCGCGCATCCGGGACACCA 3692
|||||  |||||  |||||  |||||  |||||
642 LeuAsnProAspAlaGlyGlyAlaTrpGlyArgGlyPhe..... 654
3693 ACCTACCGCTTCGCAAGATTTCCGCGCTACCGCGCAACACGACCTGC 3742
555 ..... 555
3743 GCCAATCGGTATGCAAAAAACCTCGGACGCGCGCGCTGGCATCTCG 3792
661 ..... 661
3793 TTTTCGCAACCGCGCGCAACACCTTCGACGAC..... 3828
|||||  |||||  |||||  |||||  |||||
662 .....AsnArgAlaGlyArgArgPheSerGlnLysValAlaGlyPhe 675
3829 .....GGCATCGCAACTCGGCGCGCGCTTCCCGACGCTGC.....G 3865
|||||  |||||  |||||  |||||  |||||
675 eGluLeuGlyAlaAspHisAlaValAlaValAlaGlyLysTrpTrpHis 692
3866 TTTTCGGCAATACGCGCATCGGAGGTTCGCGCATCGCATCGCATCGCGCGC 3915
|||||  |||||  |||||  |||||  |||||
692 euGlyGlyLeuAlaGlyTyrThrArgGlyAspArgGlyPheThrGlyAsp 708
3916 GCGGCTTTTACTAGCGCGACCTT..... 3939
|||||  |||||  |||||  |||||  |||||

```

```

709 GlyGlyGlyHisThrAspSerValHisValGlyGlyTyrAlaThrTyrTrl 725
3939 ..... 3939
725 eAlaAspSerGlyPheTyrLeuAspAlaThrLeuArgAlaSerArgLeuG 742
3940 .....TCAGCGCG.....ATCAGAGCAAA 3960
742 LuAsnAspPheLysValAlaGlySerAspGlyTyrAlaValAlaGlyLys 758
3961 ATCCGCCCGCGCGCTGCTATTCAGCATTCAGCAAGATACCGCGAG 4010
|||||  |||||  |||||  |||||  |||||
759 TyrArgThr.....HisGlyValGlyAlaSerLeuGlnAlaGly 771
4011 T.....TTGGCGGATTCGCGCATTCAGCAACCGCACATCGCGC 4045
|||||  |||||  |||||  |||||  |||||
771 YArgArgPheThrHisAlaAspGlyTyrPheLeuGlnProGlnAlaGly 788
4046 CAACGGCGCTATTTGTCGCAAAAGCGGATTCGCGATACGAAAC...GTC 4092
|||||  |||||  |||||  |||||  |||||
788 euAlaValPheArgAlaGlyGlyGlyAlaTyrArgAlaAlaAsnGlyLeu 804
4093 AATATCGCCACCGCGCGCTTCGATTCACCGCTACCGCGCGCGCATTA 4142
|||||  |||||  |||||  |||||  |||||
805 ArgValArgAspGlyGlyGlySerSerValLeuGlyArgLeuGlyLeuG 821
4143 GCGGATTTATTCATTCAACCGCGCGCAACACATTTCATCAGCGCTATT 4192
|||||  |||||  |||||  |||||  |||||
821 uValGlyLysArgIleGlyLeuAlaGlyArgGlnValGlnProTyrTrl 838
4193 TGACCGTCTCTATACGATCCGCGCTTCGCGCAAGTCGCAACCGCGCTC 4242
|||||  |||||  |||||  |||||  |||||
838 IeLysAlaSerValLeuGlnGlnPheAspGlyAlaGlyThrValHisThr 854
4243 AATACCGCGCTATTTGCGCGAGATTTTCCGCAAAACCGCGAGTGGATG 4292
|||||  |||||  |||||  |||||  |||||
855 AsnGlyLeuAlaHisArgThrGlnLeuArgGlyThrArg...AlaGlyLe 870
4293 GGGCGTAAAGCGCGAATCAAGTTTCAGCGTCTGCTCCAGCGTCCG 4342
|||||  |||||  |||||  |||||  |||||
870 uGly.....LeuGlyMetAlaAlaAla 878
4343 CCGCGCAAGGGCGCGCAATTTGAAAGC.....CAGCAGACGCGGCGATC 4386
|||||  |||||  |||||  |||||  |||||
878 euGlyArgGlyHisSerLeuTyrAlaSerTyrGlyLysTrpGlyPro 894
4387 AATTAGCTACCGCTGC 4404
|||||  |||||  |||||  |||||  |||||
895 LysLeuAlaMetProTrp 900
seq_name: SwissProt_40:OMP_B_RICTY
seq_documentation_block:
ID OMP_B_RICTY STANDARD: PRT; 1645 AA.
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (OMPb)
DE (OMP B) (Contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMP OR SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON.
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline

```



RT surface layer protein of Rickettsia typhi.";  
 RL Gene 133:129-133(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC STRAIN=WILMINGTON;  
 RX MEDLINE=92114896; PubMed=1370573;  
 RA Ching W.M., Carl M., Dasch G.A.;  
 RT Mapping of monoclonal antibody binding sites on CNBR fragments of  
 the S-layer protein antigens of Rickettsia typhi and Rickettsia  
 prowazekii.";  
 RL Mol. Immunol. 29:95-105(1992).  
 RN [3]  
 RP IDENTIFICATION OF CLEAVAGE SITE.  
 RX Hackstadt T., Messer R., Cieplik W., Peacock M.G.;  
 RT "Evidence for proteolytic cleavage of the 120-kilodalton outer  
 membrane protein of rickettsiae: identification of an avirulent  
 mutant deficient in processing.";  
 RL Infect. Immun. 60:159-165(1992).  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: L04661; AAB48987.1.  
 DR InterPro: IPR003858; romPA\_romPB.  
 DR Pfam: PF02708; romPA\_romPB; 1.  
 KM Antigen: S-layer; Transmembrane; Cell wall.  
 FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT TRANSMEM 1354 1645 32 KDA BETA PEPTIDE.  
 FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).  
 FT CONFLICT 657 657 H -> N (IN REF. 2).  
 FT CONFLICT 842 842 V -> I (IN REF. 2).  
 FT CONFLICT 1071 1071 G -> A (IN REF. 2).  
 FT CONFLICT 1306 1306 G -> S (IN REF. 2).  
 SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185DE CRC64;

alignment\_scores:  
 Quality: 267.50 Length: 1683  
 Ratio: 0.350 Gaps: 88  
 Percent Similarity: 45.455 Percent Identity: 19.311

alignment\_block:  
 US-09-303-518D-653 x OMPB\_RICTY ..  
 Align seg 1/1 to: OMPB\_RICTY from: 1 to: 1645

```

190 TTTGAGTCGGGGCGAAGATATTAGTTTACAACAAAGAGGAGATT 239
    |||:|||||
125 PHERITILEGLYALA..... 129
240 GGTGCGCAATGATGACGAAAGCCCGCATGATTGTTTCTGTGAT 289
    |||:|||||
130 ...Glylysserleuthr.....Ilethrglyhscglylet 141
290 CGCGTAAGCGCTGGCGCATTTGGCGGCGATCAATATATGTGACG 339
    |||:|||||
141 hralaglnalalathrThrlysserAlaglnasvalvalserlys 157
340 GCACATTAACGGC.....GGCTATAACAA 362
    |||:|||||
  
```

```

158 valasnalaglYAlaAlaIleasnaspasnaspleSerGlyValGlySe 174
363 TGTGATTTTGGTGGCGAGGAGAC.....AATC 391
    |||:|||||
174 rliaspheThrAlaAlaProSerValleuclupheasnleuIleasnP 191
392 CGCATGACGACCGCTTTCTTACCAAAATGTGAAAAAATAATTATAA 441
    |||:|||||
191 rothrThrclnglnuAlaProleuThrleuGlyAspAsnalalysIleVal 207
442 GCAGGACCTAACGGCCATCTTATGCGGCATTTTCATATCCGCGTT 491
    |||:|||||
208 AsnGlyAlaAsnGly.Ileleu..... 214
492 GCACAATTTGTACAGATGACAGAACCTGTGAGATGACCACTATATAG 541
    |||:|||||
215 ....AsnIleThrAsn.....GlyPheVallys 222
542 ANGGGTGAATAATCGCTGATTAATAATATCCCTGATCGTTGAAATC 591
    |||:|||||
223 ValSerAspLystrPheAlaGlyIleIysThrIleAsnIle..... 236
592 GAGACGACGACGACCAATATGCGGCTGTGTAAGAACCAATATACCG 641
    |||:|||||
237 .....GlyAspAsnGlnGlyLeuMetPheasnThrProaspAla 251
642 CGAAATTCATATCATATTTGCAAGCGCATATTTCTGGCTGCTGGTGGCA 691
    |||:|||||
251 la..AsnalaleuAsnleuGlnGly.....GlyGlyA 261
692 ATACCTTTGCACAAATGGA...TCAGGTGGTGGCACAGTCAACTTAGT 738
    |||:|||||
261 snThrIleasnPheasnGlyArgAspGlyThrGlyLeuValleuVal 277
739 AGCGAAAAAATTAACATAGCCCATATGTTTTTTTACCAACGAGAGCTC 788
    |||:|||||
278 SerLyAsnGlnGlyAsnalathrclupheasn.....ValThrGlySe 291
789 ATTTGGCGAGAGTGGCTCACCAATGTTTATCATGAT.....GCCCAA 832
    |||:|||||
291 rleuGlnGlyAsnleuGlyValIleGlnPheThrThrAlaAla 308
833 AGCAAAAGTGTAAATTAATGGGTATTCGCAACAGCAACCCCTATATA 882
    |||:|||||
308 laGlyIysLeuIleAlaAsnGly.....GlyAlaIleAsnalAlaValIle 322
883 GGAAGAAAGCAATGGCTCCAGCTAGTTCGTAAGATTGGTCTATGATGA 932
    |||:|||||
323 GlyThrAspAsnGly.....AlaGlyArgAlaIleAlaGlyPheIleValSe 337
933 AATCTTTGCTGAGATACCATTCAGTATTTTCAGACACCATCAAAATG 982
    |||:|||||
337 rValAspAsnGlyAsnalAlaIleThrIle.....SerG 348
983 GGAATACCTTTTAAAGAC.....AATATATAGCGGACGAGA 1020
    |||:|||||
348 lyeIlnVallyrAlaIysAspIleValIleGlnSerAlaAsnalaglyGly 364
1021 AAAATGCATCCCAACATTAACACTATCTCTACCTTATAGATTAANAAC 1070
    |||:|||||
365 GlnValThrPheGlnHis..... 370
1071 ACGACACCGTTCATTTGTTTATGTTCTTTATCCGACAGACAGAGAAC 1120
    |||:|||||
371 .....LeuValAspValIleLeu..... 376
1121 CTGTTATCATGCTGACAGTGGCGTCAACAGATATGA.....CCC 1161
    |||:|||||
377 .....GlyGlyLystrAsnheLystrAlaAspSer 387
1162 AGACTGAATATGAGAAAAATATTTCTTATTGCAAAAGAAAGCTGA 1211
    |||:|||||
388 LysValIleIleThrGlnAsnalaserPheGlySerThrAspPheGlyAs 404
    |||:|||||
  
```

1212 ATG.....ATACTACGCAACA 1231  
:::|||||:::  
404 nleuAlaValGlnIleValValProAsnAsnLysIleLeuThrLeuGlyAsn 421  
1232 TC.....ACCAACGC.....GGCGCGGTTGTATTT 1260  
:::|||||:::  
421 heileGlyAspAlaLysAsnAsnGlyAsnThrAlaGlyValIleThrPhe 437  
1261 GAGGTAATTTTACGCTCGCTCGCTAAACACGAA..... 1296  
:::|||||:::  
438 AsnAlaAsnGlyThrIleValSerGlyAsnThrAspProAsnIleVal 454  
1297 .....ACGTGGCAAGCGCGCGCTT..... 1317  
:::|||||:::  
454 lThrAsnIleLysAlaIleGlyValGlnGlyAlaGlyIleValGlnLeu 471  
1317 ..... 1317  
471 erGlyIleHISGlyAlaGlnLeuArgLeuGlyAsnAlaGlySerIlePhe 487  
1318 CATATCAGTATGAGTACCTTACTTGGAAAGTAAACGC..... 1359  
:::|||||:::  
488 LysLeuAlaAspGlyThrValIleAsnGlyProValAsnGlnAsnProLe 504  
1360 .GTGGCAACGACCGCTGTCCAAATCGGCAAGGACGCTGCTGTT 1408  
|||:::|||||:::  
504 uValAsnAsnAlaLeuAla..... 511  
1409 AACCAAGGGGAAACCAAGGCTCGCTGACGCTGGCGCACGTAAGTC 1458  
512 .....AlaGlySerIleGlnLeu...AspGlySerAla 521  
1459 ATCTTGATCAGCAGCGCGACGATCAAGGCAAAACAGCCCTTAGTGA 1508  
|||:::|||||:::  
522 IleIleThrGlyAspIleGlyAsnGlyAlaValAlaValAlaLeuGlnAs 538  
1509 AATCGGCTTG.....GTACGCGCGACGG 1531  
|||:::|||||:::  
538 pIleThrIleAlaAsnAspAlaSerLysIleLeuThrLeuSerGlyAla 555  
1532 GGAAGGTGCACTGAATGCCGATTAATCACTTCAACCCGACAACCTTAT 1581  
:::|||||:::  
555 snIleIleGlyAlaAsnAlaGlyGly.....AlaIleHIS 566  
1582 TTCGGCTTCGCGCGCGACGTTTGATTTG.....AACGGGCA 1619  
|||:::|||||:::  
567 PheGlnAlaAsnGlyGlyThrIleGlnLeuThrSerThrGlnAsnAsn 583  
1620 TTCGCTTTCGTTCCACCGCATTCAAATACGATGACGAGCGCATGATG 1669  
:::|||||:::  
583 eleuValAspPheAspLeuAspValThrThrAspGlnThrGlyValVal 600  
1670 TCAACCAATCAAGACAAAGATCCACCTTACCTTACAGGCAATATA 1719  
:::|||||:::  
600 spAlaSerSerLeuThrAsnAsnGlnThrLeuThrIleAsnGlySer... 615  
1720 GATATTACTACACCGGCATATACAC.....ACCTT 1751  
|||:::|||||:::  
616 ...IleGlyThrIleGlyAlaAsnThrLysThrLeuGlyArgPheAsnVal 631  
1752 GGATAGCAAAAA.....GAATGCTTACACGCTT 1783  
:::|||||:::  
631 lGlySerSerLysThrIleLeuAsnAlaGlyAspValAlaIleAsnGln 648  
1784 GGTTCGCGCAAGATGCA..... 1803  
|||:::|||||:::  
648 euValMetClnAsnAspGlySerValHisLeuThrHisAsnThrTyLeu 664  
1804 ...ACCAAAAG.....AAGGCGGCTCATCTGAATTA 1835  
|||:::|||||:::  
665 lIleThrLysThrIleAsnAlaAlaAsnGlnGlyLysIleIleValAlaAl 681  
1836 CCACCGGAGAGAGCGGATCGCACTTTACTGCTTCGCGGAGCAAAAT 1885  
:::|||||:::  
681 aaSPProIleAsnThrAspThrAla.....LeuAlaAspGlyThrAsn 696  
1886 TA..... 1887  
696 euGlySerAlaGlySerProLeuSerAsnIleHisPheAlaThrLysAla 712  
1888 ...AACGC..... 1893  
|||:::|||||:::  
713 AlaAsnGlyAspSerIleLeuHisIleGlyLysGlyValAsnLeuTyAl 729  
1894 ...ATATCAGCAACAAAC.....GCCAACTGTTTTCACGC 1930  
|||:::|||||:::  
729 asnAsnIleThrThrAspAlaAsnValGlySerLeuHisPheArg 746  
1931 GCAGACCGCACCGCACCGCTTCAATTCATTTAGAAAGCGGCTGCAAA 1980  
|||:::|||||:::  
746 erGlyGlyThrSerIleValSerGlyThrValGlyGly..... 758  
1981 ATGGAAGTATCCACACAGAGAAATCGTGGCAACAGATTGATC.. 2028  
:::|||||:::  
759 GlnGlnGlyLeuLysLeuAsnAsnLeuIleLeuAspAsnGlyThrThrVa 775  
2029 .....GACCGACATTTAAAGCGGAAACTTCATATTCAG 2065  
|||:::|||||:::  
775 lLysPheLeuGlyAspIleThrPhe.....Asn 785  
2066 GCGGCAAGCGGTGCTTCGCGCAATGTTCCAAAGTGAAGGCGATGG 2115  
|||:::|||||:::  
785 lGlyThrLysIleGlnGlyLysSerIleLeuGlnIleSerSerAsn... 800  
2116 CATTTAACCAATCAGCCCAAGCACTTTCGCTGTCGACCGCATCAAG 2165  
|||:::|||||:::  
801 TyThrIleThrAspHisIleGlySer..... 808  
2166 CCACCAATCTGTACACGCTGACGCTGACGCTGACAGTTGTACCG 2215  
|||:::|||||:::  
809 .....AlaAspAsnThrGly.....ThrL 815  
2216 AAAAACCATTACCGACGATTAAGTATGCTTATGACGAAGACGAC 2265  
:::|||||:::  
815 euGlnPheValAsnThrAspProIleThrValThrLeuAsnGlyGln 831  
2266 ATC.....AGAGCATGT 2279  
|||:::|||||:::  
832 AlaTyThrPheGlyValLeuLysGlnValMetValSerGlyProGlyAsn 848  
2280 CAGCCTTGCCGAT.....CAGCCTCATTTTAATCTCA 2311  
|||:::|||||:::  
848 eAlaPheAsnGlnIleGlyAsnGlyValAlaHisAla...IleAlaVal 864  
2312 CAGACCTTGCCACACTCAGCGCAATCTTACTGCA..... 2346  
|||:::|||||:::  
864 spSerIleSerPheGlnAsnAlaSerLeuGlyAlaSerLeuPheLeu 880  
2347 ...GGCGAGACAGCACACTATACGTTACGCGACGCGCAACCCCAAAACG 2393  
|||:::|||||:::  
881 SerLysThrProLeuAspValLeuThrIleLysSerThrValGlnGln 897  
2394 CAACCTC.....A 2401  
|||:::|||||:::  
897 TyThrValAspAsnPheAsnAlaProIleLeuValIleSerGlyLLeAsp 914  
2402 GCTCTGTGGGCAATGCCCAAGCAACATTTAATACGCAATTAACGC 2451  
|||:::|||||:::  
914 erMetIleAsnAsnGlyGlnValIleGlyAspGlnLysAsnIleIleAla 930  
2452 AACACATCGCTTCGACAAATGCTTCAATTAATCAAGCAACACCGCT 2501  
:::|||||:::  
931 LeuSerLeuGlySerAspAsn...SerIleThrValAsnSerAsnThrIle 946  
2502 ACAAAACGCGCAGTCTGACGCTTTCGACAAAGCTTAAGCAACGTAAGCC 2551

```

946 uTyralaGlyIleArgThrThrIleAsn..... 956
2552 ATTCCGACATCAAGCGCATGTCCTCCCTAGCCGATAAGCATATTCAT 2601
957 .....GlnGlyThrValThrIleu..... 962
2602 TTGAAAAACAGCCGCTTACCGGAAAAATACAGCGCGGCAAGATACG.. 2649
963 .....SerGlyGlyIleProAsnAspProGlyThrIle 973
2650 ....GCATTACACTTAAGACAGCGAATGAGCGTCGTCGGGACG.. 2694
973 eTyrglyLeuGlyLeuGlnAsnGlyAspProIleuLysGlnValThr 990
2695 .....GAATTAGCAATTTAAACCTTGACAAAGCCACC 2727
990 heThrThrAspTyraAsnAsnLeuGlySerIleIleIaThrAsnValThr 1006
2728 ATTACACTCAATTCGCCCTATCGACAGATGCGGAGCGGCAACCGG 2777
1007 Ile.....AsnAspAspValThrLeuThrIleGly 1016
2778 CAGTGGCGCAGATGCGCGCGCGCGCTTCGCGCTTCCCTATTATCG 2827
1016 yGlyIleAlaGly..... 1020
2828 TTACGCGCGCAACTTCGGCAGATCCGTTTCAACAGCGTGACGTTAAC 2877
1021 .....ThrAspPheAspGlyLysIleThrLeu 1029
2878 GCGAATATGACGTCGAGGAAACATCCGTTTATGCGAATCTTCGG 2927
1030 GlySerIleAsnGlyAsnAlaAsnValLysPheValAspArgThrPhe 1046
2928 TTACCGCGCGCAATTTGAAGTGGCGGAAAGTTCGAAAGCATTTACA 2977
1046 rHisProThrSerMetIleValSerThrLysAlaAsnGlnIleThrVal 1063
2978 CC.....TTGGCTGTCAACATACCGGCAACGAA.....CC 3009
1063 hrTyrlLeuGlyAsnAlaLeuValIleGlySerSerAspIleThr 1079
3010 GTAACTCTCGAGCAATGACGCTAGTAGAGAAACAAACACACACCGCT 3059
1080 ValAlaSerValArgPheThr.....GlyAsnAspSer..... 1090
3060 GTCCGAAATCTTAATTTCCACCTGCAAAACGACACGTCGATGCGCG 3109
1091 .GlyValGlyLeuGlnGlyAsnIleHisSerGlnAsnIleAspPheGly 1107
3110 CATGGCGTTATCAGCTATCCGCAAGAC..... 3138
1107 hrTyraAsnLeuThrIleLeuAsnSerAspValIleLeuGlyGlyThr 1123
3139 .....GGCAGTTCCGCTG..... 3153
1124 ThrAlaIleAsnGlyGluIleAspLeuThrAsnAsnLeuIlePheAl 1140
3154 .....CATATCCGCTCAAAAGACAGAGCTT 3181
1140 aaSnGlyThrSerThrTyrglyAsnAsnThrSerLeuSerThrThrLeuAl 1157
3182 CCGCAAACTCGCAAGCGGAGAAACAGAGCGCTTCGACGCGCAAA 3231
1157 snValSerAsnGlyAsnValGlyGln.....IleValIleAlaGlu 1170
3232 CAGGCACAACCTTCCGCGCAACACAGCGCGAA.....AA 3266
1171 GlyAlaGlnValAsnAlaThrThrThrGlyThrThrIleLysIleGly 1187
3267 AGACAAAGCGCAAGC.....CTTGACGGCGTGATGG 3298

```

```

1187 naSpasnaIaAsnAlaAsnPheSerGlyThrGlnThrTyThrIleuIleG 1204
3299 CCGCGCGCGCC.....AATGCCACCGAAAGCGAAGAGTTCGCCGA 3342
1204 lngIleGlyAlaArgPheAsnGlyThr.....LeuGlyAla 1215
3343 CCGCGCGCGCGCAGCGGAGGAAAAATCGCATTTATGACGCGGAGGA 3392
1216 ProAsnPheAspValThrGlyAsnAsnIlePheVal.....LysTy 1229
3393 AGAGAAAAAACGGTGCAGCGGATTAAGACCGCTTCGCGAAACAGC 3442
1229 rGluLeuIleArg...AspAlaAsnGlnAspTyraValLeuThrArgThrAla 1245
3443 CCGAAGCGGAAACCGCGCGCTACACCGCTTCGCCGCGCGCGCGCGC 3492
1245 snAspAla...LeuAsnValValThrThrAlaValGlyAsnSerAlaIle 1260
3493 CCGCGCGCGGATTTGCGCAACCGCGCAACCGCAACCGCAACCGCA 3542
1261 AlaAsn.....AlaProGlyValIle 1267
3543 GCGCGACCTGATCAGCCGCTTATGCCAATAGCGCTTGATGTAATTCGG 3592
1267 scIAsnIle.....AlaIleCysLeuGlnSerThrAspIleAla 1280
3593 CCACGCTCAACAGCGTTTCGCCGTACAGACGAATTGGAC..... 3633
1280 IaIaIaTyraAsnMetLeuLeuAlaLysAspSerSerAspValAlaThr 1296
3634 .....CGCGTGTTCGCAAGACCGCGCAACCGCGTTTGGACAAAGCG 3677
1297 PheIleGlyAlaIleAlaThrAspThrGlyAlaAlaValAlaThrValAs 1313
3678 CATCCGGGACACCAACACTACCGGTCCGAGATTTCCGCGTACCGCC 3727
1313 leuAsnAspThrGln.....LysThrGlnAspLeuGlyAsnArgL 1328
3728 AACAAACCGACCTGCCCAATCGTATGCAGAA.....AACCTGCGC 3771
1328 euGlyAla...LeuArgTyrlSerAsnSerGlnThrAlaAspIleGly 1343
3772 AGCGGCGCGCTCGCATCCTGTTTGGCACACCGGACCGGAACACCTT 3821
1344 GlySerGlnThrGlyAlaValSerSerGlyAspGluAlaIleAspGlnVa 1360
3822 CGACGACGCGATC..... 3834
1360 lSerTyrlGlyValThrAlaLysProPheTyraIleAlaGluGlnAspL 1377
3835 .....GCCACTCGGCACGCTTCGCCACGTCGCCCTT 3867
1377 yLysGlyGlyLeuAlaGlyTyrlLysAlaLysThrAla...GlyValVal 1392
3868 TTGGGCAATATAGCGCATCGCAGAGTTGACATGATGATCAGCGCGCGCC 3917
1393 ValGlyLeuAspThrLeuAlaAsnAspAsnLeuMetIleGlyAlaAlaIle 1409
3918 GGGT.....TTTATGACGCGGACCC 3937
1409 eGlyIleThrLysThrAspIleLysHisGlnAspTyrlLysGlyAspL 1426
3938 TTTTACAGC.....GGCATCAGAGCGCAAAATCCGCGCGCGTGCAT 3981
1426 yStrThrAspIleLysGlyLeuSerPheSerLeu..... 1436
3982 TAGCGATTCAGCGCAATACCGCGCAGTTCGCGC...GGATTGCGCAT 4028
1437 TyrlGlyAlaGlnGlnLeuValLysAsnPhePheAlaGlnIleSerAlaIle 1453
4029 C.....GAACGACATCGCGCAACGCGCTATTGCTC.... 4062
1453 ePheThrLeuAsnLysValLysSerLysSerGlnArgTyrlPhePheAspAla 1470

```

4063 .....CAAAAGCGATACCGATCGAAGATC 4092  
 1470 LaaenGlyLysMetLAsnLysGlnIleAlaGlyAsnTyrAspAsnIle 1486  
 4093 AATATCGCCACCCGGCGCTTGATCAACCGCTACCGCGGGCATTA 4142  
 1487 Thr.....PheGlyLysLauMe 1493  
 4143 GCGAGATTATTCATCAACCGCGCAACACATTCATCAGCGCTATT 4192  
 1493 tPheGlyTyrAspTyrAsnAlaLeuGlnGlyValLeuValThrProMe 1510  
 4193 TGAGCGTGTCTATACCGATCGCGCTCCGCAAGTCGCAAGCGCGTC 4242  
 1510 laGlyLeuSerTyrLeuLysSerSerAsnLysTyrLysGluThrGly 1526  
 4243 AATACCGCCGATTTGGCGAGATTTCGGCAA...ACCGGAGTGGCGA 4289  
 1527 ThrThrValAlaAsnLysArgIleHisSerLysPheSerAspArgIleAs 1543  
 4290 ATGGGCGGTAACGCGCAATCAAGTTTCAGCGTGTCCCTCCAC 4335  
 1543 pleuIleValGlyAlaLysValThrGlySerAlaMetAsnIleAsn 1558

seq\_name: SwissProt\_40:OMPA\_RICRI

seq\_documentation\_block:

ID OMPA\_RICRI STANDARD; PRT; 2249 AA.  
 AC P15921:  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (r ompA) (r omp A).  
 GN OMPA.  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=783;  
 RX MEDLINE=90354033; PubMed=2117568;  
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;  
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";  
 RL Infect. Immun. 58:2760-2769(1990).  
 CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M31227; AAA26380.1; -  
 DR PIR: A41477; A41477.  
 DR InterPro: IPR003858; r ompA\_r ompB.  
 DR Pfam: PF02708; r ompA\_r ompB; 1.  
 KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.  
 FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 212 286 A (TYPE I).  
 FT REPEAT 287 358 B (TYPE II).  
 FT REPEAT 359 430 C (TYPE II).  
 FT REPEAT

FT REPEAT 431 505 D (TYPE I).  
 FT REPEAT 506 577 E (TYPE II).  
 FT REPEAT 578 652 F (TYPE I).  
 FT REPEAT 653 724 G (TYPE II).  
 FT REPEAT 725 799 H (TYPE I).  
 FT REPEAT 800 874 I (TYPE I).  
 FT REPEAT 875 949 J (TYPE I).  
 FT REPEAT 950 1021 K (TYPE II).  
 FT REPEAT 1022 1093 L (TYPE II).  
 FT REPEAT 1094 1165 M (TYPE II).  
 FT REPEAT 1166 1180 TYPE I (INCOMPLETE).  
 SQ SEQUENCE 2249 AA; 224333 MW; A9D646C089DF087 CRC64;

alignment\_scores:  
 quality: 261.50 Length: 1472  
 Ratio: 0.382 Gaps: 77  
 Percent Similarity: 46.535 Percent Identity: 21.264

alignment\_block:

US-09-303-518d-653 x OMPA\_RICRI ..

Align seg 1/1 to: OMPA\_RICRI from: 1 to: 2249

655 CATATTGCAAGCGCATATTCTTGCTGCTGCT.....GGCATACC.. 696  
 938 AsnLeuAsnGlyAlaLeuSerGlnValThrGlyAspIleGlyAsnThrAs 954  
 697 ...TTTGCACAAAATGATCAGTGTGTGGCAGTCACTTAGTAGCG 742  
 954 nSerLeuAlaThrIleSerValGlyAlaGlyThrAlaThrLeuGlyGly 971  
 743 AAAAAATTAAACATAGCCCATATGTTTACCAACAGAGCTCAATT 792  
 971 laValIleLys.....AlaThrThrThrLysLeu 980  
 793 GCGCAGAGTGGCTCA.....CCAATGTTATCTATGA 824  
 981 ThrAspAlaIleSerAlaValLysPheThrAsnProValValThrGly 997  
 825 TGCCCAAAAGCAAAAGTGTAAATTAATGGG.....GTATGCAAAACG 868  
 997 yAlaIleAspAsnThrGlyAsnAlaAsnGlyIleValThrPheThrG 1014  
 869 GCAACCC.....TATATAGAAAACAAATGCTTCACCTA 906  
 1014 LysAsnSerThrValThrGlyAsnValGlyAsnThrAsnAlaLeuAlaThr 1030  
 907 GTTCGTAAAGATGTTCTATGATGAATCTTCTCTGGA..... 945  
 1031 ValAsnValGlyAlaGlyLeuLeuGlnValGlnIleGlyValLysAl 1047  
 946 .....GATACCATTCA.....GTATGTCAGCAAC 970  
 1047 AsnThrIleAsnLeuThrAspAsnAlaSerAlaValThrPheThrAsn 1064  
 971 CA.....AATATGCGCGAGAAAATCGATGCCAAC 1036  
 1097 AsnAlaLeuAlaThrValAsnValGlyAlaGlyLeuGlnValGlnG 1114  
 1037 ATAAACATATTCTCTACCTTAGATTAAGACGAACCGTTCATATG 1086  
 1114 LysGly.....ValValLysAlaAsnThrIleAsnLeu 1124  
 1087 TTAAATGTTTCTTTATTCGAGACAGAGAACTGTTATCATCTGCTC 1136

```
1125 ThrAspSnaIaSerAlaValThrPheThrAsnProValVal...ValThr 1140
      :::: :::: ||| ::::| ||| ::::
1137 AGGTGGGGTACACGTTATCGACCCAGACTGATTAATGGGAAATATTT 1186
      ::::| ::::| ::::| ::::| ::::| ::::|
1140 rGlyAlaIleAspAsn...ThrGlyAsnAlaAsnAsnGly...IleValT 1155
      ::::| ::::| ::::| ::::| ::::| ::::|
1187 CCTTATT.....GACAAGGAAAGGTGAATTG 1215
      ::::| ::::| ::::| ::::| ::::| ::::|
1155 hrPheThrGlyAsnSerThrValThrGlyAsnIleGlyAsnThrAsnAla 1171
      ::::| ::::| ::::| ::::| ::::| ::::|
1216 ATACTTACGCAACATCAACCAAGC.....GCGGGCGGTTT 1253
      ::::| ::::| ::::| ::::| ::::| ::::|
1172 LeuAlaThrValAsnValGlyAlaGlyIleThrLeuGlnAlaGlyIse 1188
      ::::| ::::| ::::| ::::| ::::| ::::|
1254 GTATTTTGAGGGTAAATTTACGGTCGCGCTAAACACGAAAGGCGG 1303
      ::::| ::::| ::::| ::::| ::::| ::::|
1188 rLeuAlaAlaAsnAsnIleAsnPheGlyAlaArgSer.....ThrLeuG 1203
      ::::| ::::| ::::| ::::| ::::| ::::|
1304 AAGCGCGGGCGGTCATATCATGTAGTGTAGCAGTACCGTTACTGGAAAGTA 1353
      ::::| ::::| ::::| ::::| ::::| ::::|
1203 LurPheAsnGlyProLeuAsnPolYgLYgLYsAlaIleProTYrTYrPhe 1219
      ::::| ::::| ::::| ::::| ::::| ::::|
1354 AACGGC...GTGGCAACGACCGGCTGTCCAAATCGGCAAGGACGCT 1400
      ::::| ::::| ::::| ::::| ::::| ::::|
1220 LysGlyAlaIleAlaAsnGlyAsnAsnAlaIleLeuAsnValAsnThrLy 1236
      ::::| ::::| ::::| ::::| ::::| ::::|
1401 GCTGTTCACAGCAAGGGGAAACCAAGCGTCGTC..... 1437
      ::::| ::::| ::::| ::::| ::::| ::::|
1236 sLeuLeuThrAlaSerHisLeuThrIleGlyThrValAlaIleGlnIleAsnI 1253
      ::::| ::::| ::::| ::::| ::::| ::::|
1438 .....ACCGTGGCGGACGGTAAA 1455
      ::::| ::::| ::::| ::::| ::::| ::::|
1253 LeGlyAlaGlyAsnLeuPheThrIleAspAlaSerValGlyAspValThr 1269
      ::::| ::::| ::::| ::::| ::::| ::::|
1456 GTCACTTAGTACGACGGGAGAGATCAAGCAAAACAAAGCCTTAG 1505
      ::::| ::::| ::::| ::::| ::::| ::::|
1270 rLeuAsnAlaGlnAsnIleAsnPheArgValaArgspSerValIleVala 1286
      ::::| ::::| ::::| ::::| ::::| ::::|
1286 rLeuSerAsnLeuThrGlyValGlyValaAsnIleLeuLeuAlaIla 1303
      ::::| ::::| ::::| ::::| ::::| ::::|
1553 ATATCAATTC...AACCCGCAACAACCTCTATTTCGGCTTCGGGGGGA 1599
      ::::| ::::| ::::| ::::| ::::| ::::|
1303 sPLeuValAlaProGlyAlaAspGlyThrValaIlePheAsnGly 1319
      ::::| ::::| ::::| ::::| ::::| ::::|
1600 CGTTGGATTGAAACGGGCATTCGCTTCCTCCACCGCATTCAAATAC 1649
      ::::| ::::| ::::| ::::| ::::| ::::|
1320 ValaAsnGlyLeuAsnValGlySerAsnValaIleGlyThrAlaArgAsnI 1336
      ::::| ::::| ::::| ::::| ::::| ::::|
1650 CGATGAAGGGGCG.....ATGATTGTCAACCA 1678
      ::::| ::::| ::::| ::::| ::::| ::::|
1336 eGlyAspGlyGlyAsnLysPheAsnThrLeuLeuIleTYrAsn... 1351
      ::::| ::::| ::::| ::::| ::::| ::::|
1679 ATCAAGACAAACATCCACCGTTACCATTCACGCAATAAAGATTAATCT 1728
      ::::| ::::| ::::| ::::| ::::| ::::|
1352 .....AlaValThrIleThrAspAspValaAsnLeuIu 1362
      ::::| ::::| ::::| ::::| ::::| ::::|
1729 ACAACCGCAAT.....ACAACAACCTGTATGCAAAAAGAAAT 1769
      ::::| ::::| ::::| ::::| ::::| ::::|
1363 GlyIleGlnAsnValLeuIleAsnLysAsnAlaAspPheThrSerSeth 1379
      ::::| ::::| ::::| ::::| ::::| ::::|
1770 TGCCCTACAAC...GGTTGGTTGGCGAGAAAGATCAACAAACG... 1812
      ::::| ::::| ::::| ::::| ::::| ::::|
1379 rAlaPheAsnAlaGlyAlaIleGlnIleAsnAspAlaThrTYrThrIleA 1396
      ::::| ::::| ::::| ::::| ::::| ::::|
1813 .....AACGGCGGCTCAATCTG..... 1830
      ::::| ::::| ::::| ::::| ::::| ::::|
1396 sPAlaAsnAsnGlyAsnLeuAsnIleProAlaGlyAsnIleGlnPheAla 1412
      ::::| ::::| ::::| ::::| ::::| ::::|
1830 ..... 1830
```

---

```
1413 HisAlaAspAlaGlnLeuValLeuGlnAsnSerSerGlyAsnAspArgTh 1429
      ::::| ::::| ::::| ::::| ::::| ::::|
1831 .....AATTACCAACCGGAAGAACGGAT..... 1854
      ::::| ::::| ::::| ::::| ::::| ::::|
1429 rIleThrLeuGlyAlaAsnIleAspProAspAsnAspArgIuGlyIleV 1446
      ::::| ::::| ::::| ::::| ::::| ::::|
1855 .....CGCACTTACTCTCTCCGGCGGA 1878
      ::::| ::::| ::::| ::::| ::::| ::::|
1446 AlIleLeuAsnSerValThrAlaGlyLysLysLeuThrIleAlaGly 1462
      ::::| ::::| ::::| ::::| ::::| ::::|
1879 ACAAAATTAACGCGCAATATCGACGCAACAAACG..... 1912
      ::::| ::::| ::::| ::::| ::::| ::::|
1463 LysThrPheGlyGlyAlaHisLysLeuGlnThrIleLeuPheLysGlyA 1479
      ::::| ::::| ::::| ::::| ::::| ::::|
1913 ..GCAAACTGTTTTCAGCGGACGACGACCGACCGCTACATCAATT 1960
      ::::| ::::| ::::| ::::| ::::| ::::|
1479 LagLYAspCysSerThrAlaGlyThrThrPheAsnThrThrAsnIleVal 1495
      ::::| ::::| ::::| ::::| ::::| ::::|
1961 TAGGAAGCGGTCGTCCAAAATGAA...GCTATCCCAAGAGAGAAATC 2007
      ::::| ::::| ::::| ::::| ::::| ::::|
1496 LeuAspIleThrGlyGln...LeuGlnLeuGlyAlaThrThrAlaAsnVal 1511
      ::::| ::::| ::::| ::::| ::::| ::::|
2008 GTGTGGACACAGATTGATCGACCGCACATTTAAAGCGGAA..... 2049
      ::::| ::::| ::::| ::::| ::::| ::::|
1512 ValLeuPheAsnAspAlaValGlnLeuThrGlnThrGlyAsnIleGly 1528
      ::::| ::::| ::::| ::::| ::::| ::::|
2050 .....AACTTCATATTACAGCGGACGACAGCGGTGTTCCCGCATG 2092
      ::::| ::::| ::::| ::::| ::::| ::::|
1528 yPheLeuAspPheAsnAlaLysAsnGlyIleValThrLeuAsnAsnVal 1545
      ::::| ::::| ::::| ::::| ::::| ::::|
2093 TT.....GCCAAGTGAAGCGGATGGCATTTAAGCAATCAGCC 2133
      ::::| ::::| ::::| ::::| ::::| ::::|
1545 AlAsnValaIleGlyAlaValGlnAsnThrGlyThrAsnAsnGlyThr 1561
      ::::| ::::| ::::| ::::| ::::| ::::|
2134 CAAGCAGTTTTCGTTGTCGCCACCGCATCAACCCACACATGTGACAG 2183
      ::::| ::::| ::::| ::::| ::::| ::::|
1562 LeuIleValLeuGlyAlaSerAsn.....LeuAsnAr 1572
      ::::| ::::| ::::| ::::| ::::| ::::|
2184 TTCGCACTGAGCGGTCGTGCAAGATTGTACCGMAAAACATTACGACG 2233
      ::::| ::::| ::::| ::::| ::::| ::::|
1572 yValaAsnGlyIleAlaIleMetLeuLysValaIleGlyAsnValThr... 1587
      ::::| ::::| ::::| ::::| ::::| ::::|
2234 ATAAAGTGAATGCTTCATTCAGACAGACCGACATC..... 2268
      ::::| ::::| ::::| ::::| ::::| ::::|
1588 .....IleAlaLysGlyGlyLysValaIleGlyIleGlnGly 1601
      ::::| ::::| ::::| ::::| ::::| ::::|
2269 AGAGCAATGTCAGCCTTGCCGATCAGCGTCATTAATCTCAGAGACT 2318
      ::::| ::::| ::::| ::::| ::::| ::::|
1602 ThrGlyThrAsnThrLeuThrLeuProAlaHisPheAsnLeuThrGly 1618
      ::::| ::::| ::::| ::::| ::::| ::::|
2319 TGCACACCTCAACGCG.....AATCTTAGTCAGGCG 2350
      ::::| ::::| ::::| ::::| ::::| ::::|
1618 rIleAsnLysThrGlyGlyGlnAlaLeuLysLeuAsnPheMetAsnGly 1635
      ::::| ::::| ::::| ::::| ::::| ::::|
2351 GAGACACGCACTATACGTTACGCGACGACGCCAACCAACGCAACCTC 2400
      ::::| ::::| ::::| ::::| ::::| ::::|
1635 LysValaSerGlyValaIleGlyThrAlaAlaAsnSerValaGlyAspIle 1651
      ::::| ::::| ::::| ::::| ::::| ::::|
2401 AGCCTGCTGGGC.....AATGCCCAAGCAAC 2426
      ::::| ::::| ::::| ::::| ::::| ::::|
1652 ThrThrAlaGlyAlaThrSerPheAlaSerSerValaAsnAlaLysGlyTh 1668
      ::::| ::::| ::::| ::::| ::::| ::::|
2427 ATTTAATCAAGCCATTAACGCAACACATCGCTGCGCAAAAGCTT 2476
      ::::| ::::| ::::| ::::| ::::| ::::|
1668 F.....AlaThrLeuGlyGlyThrThrSer..... 1676
      ::::| ::::| ::::| ::::| ::::| ::::|
2477 CATTTAATCTAAGCAACACGCGCTACAAAAGGAGCTGACGCTTCC 2526
      ::::| ::::| ::::| ::::| ::::| ::::|
1677 .....PheAlaAsnThrPheThrAsnThrGlyAlaValaThrLeu... 1689
      ::::| ::::| ::::| ::::| ::::| ::::|
2527 GACAACGCTAAGCAACAGTAAGCATTCGCGACTACAGCGCAATGTC 2576
      ::::| ::::| ::::| ::::| ::::| ::::|
1690 .....AlaLysGlySerIleThrSerPheAlaLysAsnValThrAlaTh 1704
      ::::| ::::| ::::| ::::| ::::| ::::|
```

```
2577 C.....CTAGCGATAGCAGATATTCATTGTAAGAACAGC...CGCT 2617
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1704 rSerPheValAlaAsnSerAlaThrIleAsnPheSerAsnSerLeuAlaP 1721
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2618 TTRACGGAAAAAATCAAGCGCGCAAGATAGCGATTACACTTAAAGAC 2667
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1721 heAsnSerAsnIleThrIleGlyGly..... 1728
2668 AGCGAATGACGCTGCCCTCGGGGCGGAATTAGCAATTAAACCTTGA 2717
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1729 .....GlyThrThrLeu...ThrLeuGlyAlaAs 1737
2718 CAACGCCACCATTTACATCAATTCGCCCTATGACACAGATGGGAGGCG 2767
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1737 ngInValThrTyThr..... 1742
2768 CGCAAACCGCGAGTGGCGAGATGCGCGCGCGCGCTTCGCGCGCTTC 2817
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1743 ..GlyThrGlySerPheThrAspThr.....Leu 1751
2818 CTATTATCGGTACGCGCGCAACTTCGGGAGATCCCGTTTCAACAGCT 2867
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1752 ThrLeuAsnThrThrPheAspGlyAlaAlaTySerGlyGlyAsnIleLe 1768
2868 GACGGTAAACGGCAAA.....TTGAAACGGTCAGGGAACATTCGCGT 2908
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 uIleTySerGlySerThrLeuAspLeuSerGlyValSerThrLeuAlaL 1785
2909 TTATGTCGGAACCTTCGGCTACCGACCGCAAAATGAACTGGCGGAA 2958
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1785 euValValThrAlaThrAsnPheAspMetAsnAsnIle..... 1797
2959 AGTTCGAGGACGACTTACACCTTGCGTGCACAAATACC.....GGCA 3002
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1798 SerProAspThrTyThrValIleSerAlaGluThrAlaGlyGlyLe 1814
3003 CGAACCCGTAAGTCTCGACCAATTCAGCTAGTGAAGAAAGAACACA 3052
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1814 uIysProThrSerTySerGluAsnValLysIleThrIleAsnAsnAspAsn 1831
3053 CACCGCTGCC.....GAAATCTTAATTCACCCCTCGAAAGCAA 3093
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1831 rgPheValAspPheThrPheAspAlaSerThrLeuThrLeuPheAlaL 1847
3094 CACGTGATGCGGCGCATGGCGTATTCAGCTATCCGAAAGAGCGGGA 3143
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1848 AspIleAlaIleAspVal.....IleAspGlyAs 1857
3144 GTTCGCGCTGCATTAATTCGGTCAAGAAAGAGCTTCGCAAAATCG 3193
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1857 pPheAlaProGlyGlyProLeu..... 1864
3194 GCAAGCGCGGAGAAACAGAGCGCGCTTCAGCGCAAAACAGGCAACTT 3243
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1865 ....AlaAsnIleProAsnAlaIleAsnIleTyLysSerLeuGluLeu 1879
3244 GCGCGCAAAACACAGCGGAAAAAGACAACGCGCAAGCCTTGAGCGCT 3293
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1879 ..... 1879
3294 GATTGCGCGCGCGCAATGCCACGAAAGAGCAAAAGTGTGGCGAAC 3343
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1880 .....MetGluAspAlaProAsnGlySerAsp. 1888
3344 GCGCGCGGAGGCGAGCGGGAATTCGCCGATATTCAGCGCGAGGAA 3393
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1889 ..AlaArgGluAla...PheAsnAsnPheGlyLeuMetThrProLeuGln 1903
3394 GAGAAAAAAGCGGTCAGCGGATTAAGACACCGCGCTTGCGGAAACAGCG 3443
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1904 Glu.....AlaAspAlaThrThrIleLeu..... 1912
3444 CGAAGCGAAACCCGCGCGCTACACCGCCTTCCCCCGCGCGCGCG 3493
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1912 ..... 1912
3494 CCCCGCGGATTTGCCGCAACCGAGCCGCCAACCGCAACCGCAACGCGAG 3543
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1913 .....GlnAspValIleValTyProSer..... 1919
3544 CGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTCCG 3593
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1920 ..AspThrIleAlaIleAlaValAsnAsnGlnValAla.....SerAs 1933
3594 CACGCTCAACAGCGTTTTCGCCGACGAGCAAGATTCAGCGCGCTTTTG 3643
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1933 nIleSerSerAsnIleThrAlaLeuAsnAlaArgMetAspLysValGln 1950
3644 CC.....GAAGACCGCGCC 3657
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1950 ergLysnIysGlyProValSerSerGlyAspGluAspMetAspAlaLys 1966
3658 AACCGCGTTTGACACAGCGC...ATCCGGAGACCAACAAACATACGCTTC 3704
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1967 PheGlyAlaThrPheSerProPheValGlyAsnAlaThrGlnIysMetCy 1983
3705 GCAAGATTTCCGCGCTACCGCCCAACCAACGACCTCGCCCAATCGCT. 3753
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1983 sAsnSerIleSerGlyTyTySerAspThrThrGlyGlyThrIleGlyP 2000
3754 .....ATGCAGAAAAACCTCGGACGCGG..... 3777
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2000 heAspGlyPheValSerAspAspLeuAlaLeuGlyLeuAlaTyThrArg 2016
3778 ..CGCGTCGCATCTGTTTCGCAACACCGGAGCGGAACACCTTCGA 3824
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2017 AlaAspThrAspIleTyLysLeuLysAsnAsnLysThrIleLysPlys..... 2031
3825 CGACGCGATCGCCCAACTCGCGACGCGCTTGCCCAAGCGCGCTTCGCGC 3874
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2032 .....AsnIysValGluSerAsnIleTySerLeuTyGlyL 2044
3875 AATAGGCGATCGGCGAGGTTGCACATCGGCATCGCGCGGCGGCTTTT 3924
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2044 euTyAsnValProTyGluAsnLeuPheValGluAlaIleAlaSerTy 2060
3925 AGTAGCGGACGCTTCAGACGATCAGAGGCAAAATCCGCGCGCGCT 3974
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2061 SerAspAsnLys.....IleArgSerTySerArgArgValI 2073
3975 GCTGATTAAGGCAATTCAGCAAGATACCGCGAGGTTTCGCGGATTCG 4024
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2073 eAlaThrThrLeuGluThrValGlyTyTyGlnThrAlaAsnGlyLysTyL 2090
4025 GCATCGAACCGCACATCGCGCAACGCGTATTTGCTCAAAAAGCGAT 4074
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2090 ysserGluSerTyThrGlyGlnLeu.....MetAlaGly 2101
4075 TACCGATAC.....GAAACGTCAATATGCCACCGCGCGCTTCG 4115
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2102 TyThrTyThrMetProGluAsnIleAsnLeu...ThrPro...LeuAl 2116
4116 ATTCAACCGCTACCGCGCGGCGCATTAAGGAGCATTAATTCATAACCGG 4165
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2116 agLyLeuArgTy...SerThrIleLys...AspLysGlyTyLysGlu 2131
4166 CG.....CAACACATTTCCATTCACGCGCTTAATTTGACCGCTGCC 4203
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2131 hrGlyThrThrTyGlnAsnLeuThrValLysGlyLysAsnTyAsnThr 2147
4204 TATACGATGCGCGCTCGGCAAGTCCGAACGCGGTCATAATTCAC. 4248
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2148 PheAspLysLeuLeuGlyAlaLysValSerSerAsnIleAsnValGln 2164
4249 .....GCCGTATGCGCGCAGAGATTTCGCA 4273
```

```

1164 uilleValLeuThrProGluLeuTyrAlaMetValAspTyrAlaPheTysA 2181
4274 AAACCGCGATGCGGATGGCGGTAACCGCGAATCAAGCTTTCACG 4323
2181 snlyValSerAla.....IleAspAlaArgLeuGlnIlyMetThr 2194
4324 CAGTCCCTCCACGCTGCCGCGCAAGGGCGCCGCAATTGGAAGCGCACGA 4373
2195 AlaProLeuProThrAsnSerPheTys.....GlnSerIlyThrSerP 2209
4374 CAGCGCGGGCATC 4386
2209 eaSpValGlyVal 2213

seq_name: SwissProt_40:OMPA_RICCN

seq_documentation_block:
ID OMPA_RICCN STANDARD; PRT; 2021 AA.
AC Q52657; P95591; P95592; P95594; Q52667; Q52668; Q52669;
AD Q52670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (OMP A) (rOMP A).
GN OMPA OR RCI273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_Taxid=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RL conorii (Malish 7 strain).";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barde V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rOMP A.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-1012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rOMP A.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -1- S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- PM: GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 001028; AAI17405.1; -
DR EMBL; AE008674; AAL03811.1; -
DR EMBL; U43794; AAB49549.1; -
DR EMBL; U43798; AAB49550.1; -
DR EMBL; U43806; AAB49551.1; -
DR EMBL; U45244; AAB49556.1; -
DR EMBL; U46918; AAB46663.1; -
DR EMBL; U83440; AAC35176.1; -
DR EMBL; U83443; AAC35179.1; -
DR EMBL; U83448; AAC35184.1; -
DR EMBL; U83453; AAC35189.1; -
DR InterPro: IPR003858; romPA_rOMP B.
DR Pfam: PF02708; romPA_rOMP B.1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
KW Complete proteome.
FT FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1877 1877
FT CONFLICT 10 10
FT CONFLICT 92 92
FT CONFLICT 126 126
FT CONFLICT 137 137
FT CONFLICT 157 157
FT CONFLICT 368 369
FT CONFLICT 374 388
FT CONFLICT 640 640
FT CONFLICT 669 669
FT CONFLICT 793 793
FT CONFLICT 804 804
FT CONFLICT 809 823
FT CONFLICT 898 898
FT CONFLICT 908 908
FT CONFLICT 985 985
FT CONFLICT 1009 1009
FT CONFLICT 1013 1013
FT CONFLICT 1182 1182
FT CONFLICT 1314 1314
FT CONFLICT 1451 1451
FT CONFLICT 1624 1624
FT CONFLICT 1628 1628
FT CONFLICT 1872 1872
FT CONFLICT 1875 1875
FT CONFLICT 1878 1879
FT CONFLICT 1936 1936
FT CONFLICT 1965 1970
FT CONFLICT 1997 1997
SO SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;

alignment_scores:
Quality: 250.50 Length: 1581
Ratio: 0.352 Gaps: 78
Percent Similarity: 45.035 Percent Identity: 20.304

alignment_block:
US-09-303-518d-653 x OMPA_RICCN

```



Align seg 1/1 to: OMPA\_RICCN from: 1 to: 2021

```

152 ACCAATACATATGCGCACTTGGCGAAATAAAGCAAGTTGCGAGTGG 201
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 ThrAsnAlaLeuAlaThrIleSerValGlyAlaGlyLysAlaThrLeuG 741
202 GCGAAGATATGAGTTACAAACAAAGGAGTTGGTCGCGCAATC 250
      | :|||:|||||:|||||:|||||:|||||:|||||:|||||:
741 yclValAlaValIleLysAlaThrThrThrLysLeuThrAspAlaSerA 758
251 G.....ATGACAAAGCCCGATGATTTTCT 282
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
758 lavalThrPheThrAsnProValValAlaThrGlyAlaIleAspAsnThr 774
283 GTGGATCCCGTAACGGCGTGGCGCATTTGGCGCATCAATATTTGT 332
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
775 GlyAsnAlaAsnAsnGlyIleAlaThrPheThrGlyAspSerThrValTh 791
333 G...AGCGTGCAATACGGCGCTATACAAATGTGATTTGGTGGCG 379
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
791 rGlyAsnIleGlyAsnThrAsnAlaLeuAlaThrValAsnValGlyAla. 807
380 AGGGAAGCAATCCGATCAGACCGCTTTCTTACCAATTTGAAAGA 429
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
808 .....GlyLeuLeuArgValGlnGlyGlyValValLysSer 819
430 AATTAATTATAAGCAGGCGCTAACGGCCATCTTATGGCGCGATTAICA 479
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
820 AsnThrIleAsnLeuThrAspAsnAla..... 828
480 TATGCCGCGTTTGCACAAATTTGTCACAGATCAGACACCTGTGAGATGA 529
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
829 .....SerAlaValThrPheThrAsnProValValValT 840
530 CCAGTTATATGATGGTGGGAATACGCTGTTAATAATTAATCCCTGAT 579
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
840 hGlyAlaIleAsp..... 844
580 CGTGTTCGAATCGGAGCAGCAGACATATTTGGCGTGTGATGAAGAGA 629
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
845 .....AsnThrGlyAs 848
630 ACCCAATACCGCGAAGTTCAATATTCATTCGAAGCCGATATTTGTCG 679
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
848 nAlaAsnAsnGlyIleValThrPheThrGlyAspSerThrValThrGlyA 865
680 TCGTGGGTGC...AATACCTTGCACAAATGATCAGGTGGTGGCACA 726
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
865 snIleGlyAsnThrAsnAlaLeuAlaThrIleSerValGlyAlaGlyLys 881
727 GTCAACTTGTAGCGAAATAAATTAACATACCCATATGTTTTTTACC 776
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
882 AlaThrLeuGlyGlyAlaIleLys.....Al 891
777 AACAGAGGCTCATTTGGCGACAGTGCCTCA.....C 808
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
891 aThrThrThrLysLeuThrAspAsnAlaSerAlaValThrPheThrAsn 908
809 CATGTTTATCTATGATGCC...CAAAAGCAAAAGTGGTTAATTAATGG 855
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
908 roValValValThrGlyAlaIleAspAsnThrGlyAsnAlaAsnAsnGly 924
856 GTATTGCAA.....ACAGCAACCCCTATATATAG 884
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
925 lIleValThrPheThrGlyAspSerThrValThrGlyAsn.....IleG 939
885 AAAAAACAATGGCTCCAGCTAGTTCTGAAGATTTGGTTATGATGAAA 934
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
939 yAsnThrAsnAlaLeuAlaThrVal.....AsnV 949
935 TCTTTGCTGGAGATACCCATTCAGTATTTCTACGAACCATCAAAATGGG 984
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

949 aGlyAlaGlyValThr..... 954
985 AATATCTTTTAAAGCAATATATATGCGCGAGAAATTCATGCCAA 1034
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
955 .....LeuGlnAlaGlyLysLeuAspAlaAs 964
1035 ACATTAACACATATCTCTACTTATATTAATAAAGCAAGACCGTTCAAT 1084
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
964 nasnIleAspPheGly.....AlaArgSerThrLeuG 975
1085 TGTTAATGTTTCTTATCCGAGACAGACAGAACCTGTTATCATGCT 1134
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
975 lupheAsnGlyProLeuAsp..... 981
1135 GCAGGTGGGCTCAACACT.....TATGACCCAGACTGAATAA 1172
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
982 ..GlyGlyGlyAsnAlaIleProTyrTyrPheLysGlyAlaIleAlaAs 997
1173 TGGAGAAATATTTCTTATTTGACAAAGAAAGTGATTTGATCTTA 1222
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
997 nGlyAsnAsnAlaIleLeuAsnValAsnThrLys.....LeuLeuT 1011
1223 CC.....AGCAATCAACCAAGCGCG 1245
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1011 hralaTyrHisLeuThrIleGlyThrValAlaGluIleAsnIleGlyAla 1027
1246 GCGGCTTTGTA.....TTTGAGGCTAATTTACGCTTCGCC 1283
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1028 GlyAsnLeuPheAlaIleAspAlaSerAlaGlyAspValThrIle..... 1042
1284 TAAAAACACAGAACGTGCGAAGCGCGCGCTTCATATAGT.....G 1327
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1043 .....LeuAsnAlaGlnAspIleHisPheArgAlaLeuA 1054
1328 ATGCGACTACCTTACTTGAAGTAACGCGCTGGCAACGACCGCGCTG 1377
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1054 sPserAlaLeuValLeuSerAsnLeuThrGlyAla..... 1065
1378 TCCAAATTCGCGCAAGCGCTGCTGTCA.....GCCAA 1415
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1066 .....GlyValAlaAsnAsnIleLeuLeuAlaAlaAspLeuValAlaPr 1079
1416 AGGGAACCAACGAGCTCGTGCAGCGCGCGAGCAATGTCATCTTAG 1465
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1079 oGlyValAspIleGlyThrValAlaIlePheAspGlyGlyValAsnGlyLeuA 1096
1466 ATCAGCAGCGGCGAGCATCAAGCGCAAAACACCTTACTGAATCGGC 1515
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1096 snIleGlySerAsnValAlaGlyAlaAlaArgAsnIleGlyAspValGly 1112
1516 TTGCTCAGCGGCGAGGCGAGCTGCACATGATCCGATATCATGTTCAA 1565
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 .....GlyAsnLysPheAs 1117
1566 CCCCAGCAAACTCATTTTCGGCTTTCGCGCGGACGTTGATTTGAACG 1615
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1117 nThrLeuLeuIleTyr..... 1122
1616 GGCATTGCGCTTTCGTTCCACCGCATTCAAATACCGTAGAGGGCGATG 1665
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1122 ..... 1122
1666 ATTGTCACACCAATCAAGAACAAGATCGAGTTACATTAACAGCAA 1715
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1123 .....AsnAlaValThrIleThrAspAs 1130
1716 TAAAGATATTACTACAACCGCAAT.....AACCAACACTTGATA 1756
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1130 pValAsnLeuGlnGlyIleGlnAsnValLeuIleAsnAsnAlaAsp 1147
1757 GCAAAAAAGAAATTTGCTACAC...GGTTGGTTTGGCGGAAAGATGCA 1803
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1147 heThrSerSerThrAlaPheAsnAlaGlyThrIleGlnIleAsnAspAla 1163

```

```

1804 ACCAAACG.....AACGGCGGCTCAATCTG.....AA 1832
1164 ThrTyrThrIleSpAlaAsnSnglyAsnLeuAsnIleProAlaGlyAs 1180
1833 TTAACAACGGGAAGAACGGATCGCCTTACTGCTTCCGGCGGAACA 1882
1180 nIleYsPheAlaHsIAlaSpAlaGlnIleuIleuGlnAsnSerSerg 1197
1883 ATTAAACGGCAATATCAAGCAACAACGCAACTGTTTTCAGCGGC 1932
1197 LysAsnSparYThrIleThr..... 1203
1933 AGACGACACCGACCGCTCAATCATTTAGAGCGGGTGGTCAAAAT 1982
1204 .....LeuGlyAla..... 1206
1983 GGAAGGTATCCACAAGAAATCGTGGGCAACGAT..... 2022
1207 .....AsnIleAspProAspAsnAspSgIuglyI 1217
2023 .....TGAATCGACCGCATTTAAAGCGGAAACTTCATATTCAGGCG 2067
1217 leValIleuAsnSerValThrAlaGlyLysLysLeuThrIleAlaGly 1233
2068 GGACAAAGCGGTGGTTCCCGCAAT.....GTTGCCAAAGT 2102
1234 GlyLysThrPheGlyGlyAlaHsLysLeuGlnAspIleValPheYsG 1250
2103 GGAAGCGATTGGCATTTAAGCAATCAGCCCAAGCAGTTTTCGGTGG 2152
1250 yGlyIuglyAsp.....PheGlyThrA 1257
2153 CACCGCATCAAAAGCCACAACTGTACACGTTGGCATGGACGGGTGG 2202
1257 laGlyThrThrPheAsnThrThrAsnIleValIleuAspIleThrGlyGln 1273
2203 .....ACAATTGTACGAAAAAACCATTAACGAGATTAAGTATTCG 2246
1274 LeuGlnLeuGlyAlaThrThrAlaAsnValValLeuPheYsAspAlaVa 1290
2247 TTCATTGACAAAGCCGACATCAAGAGGCAATGTC..AGCTTGGCGATC 2293
1290 LgInLeuThrGlnThr.....GlyAsnIleGlyGlyPheLeuAsp 1304
2294 AGCCTATTAAATCTCACAGGACTGCACACTCACTCAAGCAATCTTACT 2343
1304 heAsnAlaYsAsn.....GlyThrValThrLeuAsnAsnValAsn 1318
2344 GCAGGGGAGACACGACTATACGTTACGGCAACGCCCAACCAAAACGG 2393
1319 ValAlaGlyThrValYsAsnThr.....GlyGlyThrAsnAsnG 1332
2394 CAACCTCACCCTCGTGGCAATGCCCAACAACTTATCAAGCACAT 2443
1332 YThrIleuIleValLeuGlyAlaSer.....AsnLeuAsnArgV 1345
2444 TAAACGCAACACATCGGCTCGAACAATGTTCAATTTAATCTAAGCAAC 2493
1345 alaSnGly.....IleAlaMet 1350
2494 AACGCGGTCAAAAGCGAGTCTGACGCTTCCGACAAACGTAAGGCAAA 2543
1351 LeuYsValGlyAlaGlyAsnValThrIleAlaYsGlyYAsnValY 1367
2544 CGTAAGCCATTCGGCACTCAAGGCAATGTCCTCAGCCGATTAAGGCA 2593
1367 sIleGlyGluIleGlnGlyThrSgThrAsnThrLeuThrLeuProAla 1383
2594 TATTCATTTGAAAAACAGCCGCTTACGGAATAATCAAGCGCGCAAG 2643
1384 .....HisPhe.....LysLeuThrGlySerIleAsnYsThrGly 1395
2644 GATACGGCATTTACACTTAAAGACAGCAATGACGCTGCCGTGGGC.. 2691
1396 GlyAlaIleuLysLeuAsnPheLeuAsnGlyGlySerValSergIYVa 1412
2692 .....ACGGAATTAGCAATTTAAACCTTGACACGCCCA 2725
1412 lValGlyThrAlaAsnSerValGlyAspIleThrThrAlaGlyAlaT 1429
2726 CCATT..... 2730
1429 hrSerPheAlaSerSerValAsnAlaYsGlyThrAlaThrLeuGlyGly 1445
2731 ACACCTCAATTCGCGCTATCGACAGATCGCGACGCGCAACCC.... 2775
1446 ThrThrSerPheAlaHsIThrPheThrAsnThrGlyAlaValThrLeuAl 1462
2776 ...GGCAGTGGGCAATGCGCGCGCGCGCTGCGCGCTTCCCTAT 2821
1462 aLYsGlySerIleThrSerPheAlaLysAsnValThrAlaThrSerPheV 1479
2822 TATCCGTTACCGCCCAACTCGGCGCAATCCGCTTC...AACAGCTG 2868
1479 AlaAla.....AsnSerAlaThrIleAsnPheGlyAsnSerLeu 1491
2869 ACGGTAAACGGCAATTTGAACGTCAGGAAACATTCGCTTATGTGGA 2918
1492 AlaPheAsnSerAsnIleThrGlySergIYThrThrLeuThrLeu.... 1506
2919 ACTCTGGGCTACCGCAGCGCAATTTGAAGCTGGCGGAAGTTCCGAA 2968
1507 .....GlyAlaAsnGlnValThrTyrThrGlyThrG 1517
2969 GCACCTTAC.....ACCTTGGCTGTCAACAATAC..... 2997
1517 LysSerPheThrAspThrLeuThrLeuAsnThrThrPheAspGlyAlaAla 1533
2998 .....GGCAAGAACCCGTAAAGTCTCGACCAATTCAGCTGATGGA 3038
1534 LysSergGlyAsnIleLeuIleLysSergIYSerThrLeuAspLeuS 1550
3039 AGGAAGAACACACACCGCTGCCGAAT...CTTAATTCACCGCTGC 3085
1550 rGlyValSerAsnLeuAlaLeuValValThrAlaThrAsnPheSspIeTa 1567
3086 AAAACGAACAGTGTGATCGCGCATGCGCTTATCAGTTATCCGCAAA 3135
1567 snAsnIleSerProAspThr.....LysTyrThrValIle..... 1578
3136 GACGGCGAGTTCGCTGCATATCGGTCAAAAGACAGAGCTTTCGA 3185
1578 ..... 1578
3186 CAACCTCGGCAAGCGGGAGAAACAGAGGCCGCTTG.....ACGGCAA 3229
1579 .....serAlaGlnThrAlaGlyGlyLeuYsProThrProL 1591
3230 AACAGCAACAATTCGCGCAACACAGAGCGGCAAAAGAACAA 3273
1591 ySgIu.....AsnValYsIleThrIleAsnAsnAspAsnArgPhe 1604
3274 .....GGCAAAAGCTTGACGCGCTGATTTGCGGCGG 3305
1605 ValAspPheThrPheAspAlaSerThrLeuThrLeuPhe..... 1617
3306 GCGCAATGCCACGAAAGGACGAAAGTGTGGCAACCGCGCGGACG 3355
1618 .....AlaGlnSppIleAlaGlyValIleAspGlnAspPheAlaP 1632
3356 CAGCGGGGGAATAATGCCGCAATTTATGAGCGGAGAGAAAGAAAACG 3405
1632 roGlyGlyProLeuAlaAsnIleProAsnAlaAlaAsnIleYsLysSer 1648
3406 GTGCAAGCGGATTAAGACACC..... 3426

```

```

1910 LeuHrValLysGlyLysAsnTrpAsnThrPheAspGlyLeuLeuGlyAl 1926
4224 CAAGATCGGACGCGCTCATACC..... 4248
      ||||| : : : : : ||| : :
1926 alyValSerSerAsnIleAsnValAsnGluIleValLeuThrProGluL 1943
      ||| : : : : : ||| : :
4249 .....GCCGATTTGGCGCAGATTTGGCCAAACCCGACGTCGGAAATGG 4293
      ||| : : : : : ||| : :
1943 euTrpIleMetValAspTrpAlaPheLysAsnLysValSerIle..... 1957
4294 GCGGTAAACGCGCAATCAAGATTTTCAGCTGTCCCTCCACGCTGCGCG 4343
      : : : : : : : : : : : ||| : : : :
1958 ...IleAspIleArgLeuGlnIleGlyThrIleAlaProLeuProThrAsn 1973
4344 CGCCAAAGGCGCGCAATTTGAAAGCCGACACAGCGCGGCATC 4386
      : ||| : : : : : : : : : : :
1973 rpeLys.....GlnSerLysThrSerPheAspValGlyVal 1985

seq_name: SwissProt_40:AIDA_ECOLI

seq_documentation_block:
ID AIDA_ECOLI STANDARD; PRT; 1286 AA.
AC Q03155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin aida-1 precursor.
GN AIDA-1.
OS Escherichia coli.
OG Plasmid pIB6.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RA MEDLINE=9232638; PubMed=1625582;
RX Benz I., Schmidt M.A.;
RT "AIDA-1, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
CC -I- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; X65022; CAA46156.1; -.
DR PIR; S28634; S28634.
KW Cell adhesion; Signal; Outer membrane; Plasmid.
FT SIGNAL 1 49
FT CHAIN 50 ? ADHESIN AIDA-1.
FT PROPEP ? 1286
FO SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

alignment_scores:
Quality: 247.50 Length: 1470
Ratio: 0.392 Gaps: 62
Percent Similarity: 42.925 Percent Identity: 18.163

alignment_block:
US-09-303-518D-653 x AIDA_ECOLI ..

```

```

688 GCGAATACCTTTGACAAATGATCAGT.....GCTGCAC 725
|||||:||||| ||| |||||
45 GlyAsnAlaPheAlaValAsnIleSerGlyThrValSerSerGlyGly 61
726 AGTCACTAGTAGCGCAAAAATTAACATACCCATATGCTTTTAC 775
|||||:||||| ||| |||||
61 rValSerSerGlyGlyThrGlnIleValTyrSerGlyArgGly..... 75
776 CACAGAGAGCTCATTTGGCAGACGGCTGACCAATGTTATCTATGAT 825
.....:..... ||| |||||: ||| |||||: |||
76 ..AsnSerAsnAlaThrValAsnSerGlyGlyThrGlnIleValAsn 91
.....:..... ||| |||||: ||| |||||: |||
826 GCCCAAGAGCAAAAGTGTAAATTAATGGGTATGCAAAACGCAACC 875
|||: ||| ||| |||: ||| |||: ||| |||: |||
92 GlyIleYsrThr.....ThrAlaThrThrValAsnSerSerGlySer 106
|||: ||| ||| |||: ||| |||: ||| |||: |||
876 CTATATAGGA.....AAAAAGCAATG 895
|||: ||| ||| |||: ||| |||: ||| |||: |||
106 nAsnValGlyThrSerGlyAlaThrIleSerThrIleValAsnSerGly 123
GCTTCAGAGCTAGTTCGTAAGATTGTTCTATGATGAATC...TTTGT 942
|||: ||| ||| |||: ||| |||: ||| |||: |||
123 LylleGlnArgValSerSerGlyValAlaSerAlaThrAsnLeuSer 139
943 GGAGATACCATTCATATTCTACGAACCATCAAAATGGAAATCTT 992
||| ||| ||| |||: ||| |||: ||| |||: |||
140 GlyGlyAlaGlnAsnIleTyrAsnLeuGlyHisAlaSerAsnThrVal 156
||| ||| ||| |||: ||| |||: ||| |||: |||
993 TTTTAAGCACAATATATAT.....GGCGAGCAAAATCGATGCCA 1033
|||: ||| ||| |||: ||| |||: ||| |||: |||
156 ePheSerGlyAsnGlnThrIlePheSerGlyGlyIleThrAspSer 173
|||: ||| ||| |||: ||| |||: ||| |||: |||
1034 AACATAACACTATTCTCTACCTTAATGATTAATAACACAGACCTTCAA 1083
|||: ||| ||| |||: ||| |||: ||| |||: |||
173 hAsnIleSerSerGlyGlyGlnArgValSerSerGlyGlyVal... 188
|||: ||| ||| |||: ||| |||: ||| |||: |||
1084 TTGTTAATGTTTCTTTATCCGAGACAGACAGACCTGTTAT... 1128
|||: ||| ||| |||: ||| |||: ||| |||: |||
189 AlaSerAsnThrThrIleAsnSerSerGlyAlaGlnAsnIleLeuSerG 205
|||: ||| ||| |||: ||| |||: ||| |||: |||
1129 .....CATGCTGACGGTGGGCTCACAGTTAT... 1155
|||: ||| ||| |||: ||| |||: ||| |||: |||
205 uGlnGlyAlaIleSerThrHisIleSerSerGlyGlyAsnGlnIle 222
|||: ||| ||| |||: ||| |||: ||| |||: |||
1156 .....CGACC 1161
222 eAlaGlyAlaAsnAlaThrGlnThrIleValAsnSerGlyGlyPheGln 238
1162 AGACTGAATATATGA..... 1176
|||||:||||| ||| |||||
239 ArgValAsnSerGlyAlaValAlaThrGlyThrValLeuSerGlyGly 255
1177 .GAATAATTTCTTTATTTGACAAAGAAAGTGAATGATGATCTACCA 1225
|||: ||| ||| |||: ||| |||: ||| |||: |||
255 rGlnAsnValSer.....SerGlyGlySerAlaIleSerThr 268
|||: ||| ||| |||: ||| |||: ||| |||: |||
1226 GCAACATCAACCAAGCGCGGCTTTGATTTGAGGGTAATTTTCG 1275
|||: ||| ||| |||: ||| |||: ||| |||: |||
268 eValIleYrAsnSerGlyValGlnThrValPheAlaGlyAlaThrVal 284
|||: ||| ||| |||: ||| |||: ||| |||: |||
1276 GTCTGCCCTMAAACAGAAAGCTGGCAGAGCGCGGCTTCATATCAG 1325
|||: ||| ||| |||: ||| |||: ||| |||: |||
285 AspThrThrValAsnSer.....GlyGlyAsnGlnAsnIleSe 297
|||: ||| ||| |||: ||| |||: ||| |||: |||
1326 TGATGCG.....AGTACGCTTACTGGAAGTAAAGCGCGGCA 1366
|||: ||| ||| |||: ||| |||: ||| |||: |||
297 rSerGlyGlyIleValSerGlyThrThrValAsnValSerGlyThrGln 314
|||: ||| ||| |||: ||| |||: ||| |||: |||
1367 AC.....GACGCGCTGTCAAAATCGGCAAAAGCAACGCTG 1401
|||: ||| ||| |||: ||| |||: ||| |||: |||
314 snIleYrSerGlyGlySerAlaLeuSerAlaAsnIleYsgIleSerGln 330

```

```

1402 CTGCTTCAAGCCAAAGGGGA..... 1422
|||: ||| ||| |||: ||| |||: ||| |||: |||
331 IleValAsnSerGlyGlyThrAlaIleAsnThrIleValIleAspGly 347
|||: ||| ||| |||: ||| |||: ||| |||: |||
1423 .....AACCAAGCTCGGTACAGCTGGCGGCGAGGTAAAGTCA 1459
|||: ||| ||| |||: ||| |||: ||| |||: |||
347 YrYrGlnHisIleArgAsnGlyGlyIleAlaSerGly.....ThrI 361
|||: ||| ||| |||: ||| |||: ||| |||: |||
1460 TCTTGATCAGCAGCGGACGATCAAGCAAAACAAACGCTTAGTCA 1509
|||: ||| ||| |||: ||| |||: ||| |||: |||
361 leValAsnGlnSerGlyTyrValAsnIleSerSerGlyGlyTyrAla 377
|||: ||| ||| |||: ||| |||: ||| |||: |||
1510 AATCGCTTGGTCAGCGGAGCGGAGCGGTGCACATGAATGCCATATCA 1559
|||: ||| ||| |||: ||| |||: ||| |||: |||
378 SerThrIleIleAsnSerGlyGlyThrIleuArgValLeuSerAspGly 394
|||: ||| ||| |||: ||| |||: ||| |||: |||
1560 GTTCAACCCCGCAAACTCTATTTCGCTTCGCGCGGAGTTTGAT 1609
|||: ||| ||| |||: ||| |||: ||| |||: |||
394 rAlaArgGlyThrIleLeuAsn.....AsnSerGlyArgGlnAsn 408
|||: ||| ||| |||: ||| |||: ||| |||: |||
1610 TGACGGGATTCGCTTCGTCACCGCATCAAAATACGATGAAGG 1659
|||: ||| ||| |||: ||| |||: ||| |||: |||
408 aIleAsnGlyValSerTyrAsnAlaMetIleAsnThrGlyGlyAsn 424
|||: ||| ||| |||: ||| |||: ||| |||: |||
1660 GCGATGATTGTCAACCAATCAAGACAAAGATCCACCGTTACCAT 1709
|||: ||| ||| |||: ||| |||: ||| |||: |||
425 GlnTyrIleTyrSerAspGlyGlnAlaThrAlaIleValAsnThrSe 441
|||: ||| ||| |||: ||| |||: ||| |||: |||
1710 AGGCAATTAAGATATTACTACCAACGGGCANT.....AACACACT 1750
|||: ||| ||| |||: ||| |||: ||| |||: |||
441 rGlyPheGlnArgIleAsnSerGlyGlyThrAlaProValGlnAsnSer 458
|||: ||| ||| |||: ||| |||: ||| |||: |||
1751 TGGATGCAAAAAGAAATTTGCC..... 1773
|||: ||| ||| |||: ||| |||: ||| |||: |||
458 aIleValIleThrArgThrValSerSerAlaAlaIleProPheAspAla 474
|||: ||| ||| |||: ||| |||: ||| |||: |||
1774 ...TACACGGT.....TGCT 1787
|||: ||| ||| |||: ||| |||: ||| |||: |||
475 ValTyrSerGlyGlyGlnThrValTyrLeuThrPargIleTyrP 491
|||: ||| ||| |||: ||| |||: ||| |||: |||
1788 T.....GGCGAAGAAAGATG 1801
|||: ||| ||| |||: ||| |||: ||| |||: |||
491 rSerAsnPheLeuThrAlaValIlePheMetPheProGlyThrAlaSer 508
|||: ||| ||| |||: ||| |||: ||| |||: |||
1802 CAACCAAAACGAAC.....GGCGGCTCAATCTGAATTCACACCGGA 1845
|||: ||| ||| |||: ||| |||: ||| |||: |||
508 lValAsnValAsnLeuSerGlyArgLeuAsn..... 518
|||: ||| ||| |||: ||| |||: ||| |||: |||
1846 GAAGCGATGCACTTACTGCTTCGCGGGAACAATTTAAACGCGCA 1895
|||: ||| ||| |||: ||| |||: ||| |||: |||
519 .....AlaPheAlaGlyAsnValValGly 527
|||: ||| ||| |||: ||| |||: ||| |||: |||
1896 TATCAGCAAAACAAACGCAACTGTT...TTCACGGC...AGACCA 1939
|||: ||| ||| |||: ||| |||: ||| |||: |||
527 rIleLeuAsnGlnGlnGlyArgGlnIleValTyrSerGlyAlaThrAla 544
|||: ||| ||| |||: ||| |||: ||| |||: |||
1940 CACCGCACGCTCAATCATTTAGGAAGCGGAGTCAAAATGAAGG 1989
|||: ||| ||| |||: ||| |||: ||| |||: |||
544 hSerThrValIleGlyAsnAsnGlnGlyArgGlnIleValLeuSerGly 560
|||: ||| ||| |||: ||| |||: ||| |||: |||
1990 ATCCCAAGAGAAATCGTGGGACACACATTTGATCGACGCAAT 2039
|||: ||| ||| |||: ||| |||: ||| |||: |||
561 IleThrAspGlyThrValLeu..... 567
|||: ||| ||| |||: ||| |||: ||| |||: |||
2040 TAAAGCGAAACTTCATATTACGGCGGACAAAGGTTTCCCGCA 2089
|||: ||| ||| |||: ||| |||: ||| |||: |||
568 .....AsnSerGlyGlyLeuGlnAlaValSerSerGly 579
|||: ||| ||| |||: ||| |||: ||| |||: |||
2090 ATGTTGCCAAAGT.....GAGCGGCTTGCATTAAGCAAT 2127
|||: ||| ||| |||: ||| |||: ||| |||: |||
579 lYrValIleSerAlaThrValIleAsnGlnGlyGlyAlaGlnPheValTyr 595
|||: ||| ||| |||: ||| |||: ||| |||: |||
2128 CACGCCAAGCAGTTTTCGCTGTCACCGCATCAAAAGCAACAAATCTG 2177

```

```

506 AspGlyGlnValThrGlyThrAsnIleuysnGlyGlyThrIleAr 612
2178 TACAGCTTCGACCTGACGGCTGACA..... 2205
612 gValAspSerGlyAlaSerAlaIleuAsnIleAlaLeuSerSerGlyAla 629
2206 .....AGTTGTACCGMAAACCATTTACCGACGATTAAGTATTGCT 2247
629 snLeuPheThrSerThrGlyAlaThrLeuProGluLeuThrThrMetAla 645
2248 TCATTGAGC...AAGACCGACATAGAGCGCATTCACCTTGCCGATCA 2294
646 AlaLeuSerValSerGlnAsnHisAlaSerAsnIleValLeuGluAsnG 662
2295 CGCTCATTTAAATCTACA.....GGACTGCC.....ACACTCA 2329
662 yGlyLeuLeuArGlyAlaThrSerGlyGlyThrAlaThrAspThrThrVala 679
2330 ACGGC.....AATCTTAGTCAGCGGAGACAGCAGCATATACG 2367
679 snSerAlaGlyArgLeuArGlyLeuAspAspGlyGlyThrIleAsnGlyThr 695
2368 GTTACGGCGCAACGCC.....ACCCAAA 2390
696 ThrThrIleAsnAlaAspGlyIleValAlaGlyThrAsnIleGlnAsnAs 712
2391 CGGCACCTGACGCTC..... 2406
712 pGlyAsnPheIleuAsnIleuAlaGluAsnTyraSppheGluThrGlu 729
2406 ..... 2406
729 euSerGlySerGlyValLeuValIysAspAsnThrGlyIleMetThrTy 745
2407 .....GTGGCAATGCCCAACG 2423
746 AlaGlyThrLeuThrGlnAlaGlnGlyValAsnValIysAsnGlyIle 762
2424 AACATTTAATCAAGCCACATTAAGCGCAACACATCGGCTTCGGAATG 2473
762 eilePheAspSerAlaValAlaAsnAlaAspMetAlaValAsnGlnAsn 779
2474 CTTCATTTATCTAAGCAGACAGCCGCTACAAAGCGCAGCTGACGCTT 2523
779 IatyrIleAsnIleSerAspGlnAlaThrIleAsnGlySer.....Val 793
2524 TCCGACAGCCTAAGCAAAAGCTAAGCCATTCGCACTCAAGCGCAATGT 2573
794 AsnAsnAsnGlySerIleValIleAsnAsnSerIleIleAsnGlyAsnI 810
2574 CTCCCTAAGCGGATAAGCAGATATTCATTTGMAAACAGCCGCTTACCG 2623
810 eThr.....AsnAspAlaAspLeuSerPheGlyThrAlaIys..... 822
2624 GAAAAATCAAGCGCGCGCAAGATACGCAATTCACCTTAAAGACAGCGAA 2673
822 ..... 822
2674 TGGACGCTCGCTCGGCGACGGAATTAGCAATTTAACTTGACAAACGC 2723
823 .....LeuLeuSerAlaThrValAsnGlySerLeu...ValAsnAsn 836
2724 CACCATTAACATCAATTCGCGCTATCGACAGATGCGGCGCGCAAA 2773
836 sAsnIleIleLeuAsn..... 841
2774 CCGGCGAGTCGGCAGATGCGCGCGCGCTTCGCGCGCTTCCTATTA 2823
842 .....ProThrIysGlnSerAlaGlyAsnThrLeu 851
2824 TCCGTTACCGCGCAACTTCGCGCAATCCGTTTCAACAGCCTGACGCT 2873

```

---

```

852 ThrValSerAsnTyThrGlyThrProGly.....SerValIleSerIle 866
2874 AAACGGMAATTAAGACGCTGACGGAACATTCGCTTATGTCGAACTCT 2923
866 uGlyGlyValLeuGluGlyAspAsnSerLeuThrAspArgLeuValVal 883
2924 TCGCTACCCGACCGCGMAATTAAGCTGGCGGAAAGTTCCGAGGACT 2973
883 ySdIysAsnThrSerGlyGln.....SerAspIleVal 893
2974 TACACCTTGCGCTCAACAATACCGCAACGCAACCCGTAAGCTCGAGA 3023
894 Tyr.....ValAsnGluAspGlySerGlyGlnThrArgAspI 907
3024 ATTGACGCTAGTGAAGAAAGACACACACCGCTGTCCGAAATCTTA 3073
907 yIleAsnIleIleSerValGluGlyAsnSer.....AspAlaG 920
3074 ATTTCACCTGCAAAAGCAACACGCTGATCCGGCGCATGCGGTATCAG 3123
920 IuPheSerLeuIysAsn...ArgValAlaIaGlyAlaTyraSptyrThr 935
3124 CTATCCGCAAGACGCGAGTTCGCTGATTAATCCGGTCAAGAACAA 3173
936 Leu..... 936
3174 AGAGCTTCCGACAAACTCGCAGAGCGGGAGAAACAGAGCGGCTTGA 3223
936 ..... 936
3224 CGCAAAACAGGACACACTTCGCGCAAAACAGCGGMAAAACACAC 3273
937 ..... 940
3274 GCGCAAAAGCTTGACGCGCTGATTCGCGCGGCGCAATGCCACGAAA 3323
940 ..... 940
3324 GGCAGAAAGTGTGCCGAACCGCGCGGACGCGGGAATAATGCGG 3373
941 .....GluSerGlyThrAspAsnIysG 948
3374 GCATTTATCGAGCGGAGAGAGAAAGAAACGGGTGCAGCGGATTAAGAC 3423
948 IYTrpTy..... 950
3424 ACCGCTTGCGAAACAGCGGAAACCGGAAACCGCGGCTACACCGC 3473
951 .....LeuThrSerHI 954
3474 CTTCGCGCGCGCGCGCGCGCGCGGATTTGCCGAACCGCAGCCCG 3523
954 sleuProThrSer.....AspThrArgGlnTyraSpproG 966
3524 AACCGCAACCCCAACCGCAGCGACCTGATCAGCCGTTATGCAATAGC 3573
966 Iu.....AsnGly 968
3574 GGTTCGATGATTTTCGCGACGCTCAACAGCCTTTTCGCGTACAGA 3623
969 SerTyraIaIaThrAsnMetAlaIleuAlaAsnSerLeuPheLeuMetAsp 985
3624 CGAATGGACCGCGGTTT.....GCCGAGACCGCC 3655
985 uAsnGluArgIysGlnPheArgAlaMetSerAspAsnThrGlnProGlu 1002
3656 GCAAGCGCGTTTG.....ACAAGCGGATCCGCGGACACAAACAC 3696
1002 eAlaSerValTrpMetIysIleThrGlyIleSerSerGlyIys... 1017
3697 TACGTTGCGAAGATTTCGCGGCTTACCGCAACAAACGACCTGGCCA 3746
1017 ..... 1017

```

```

3747 AATCGTATGCAGAAAAACCTCGGCACGCGCGCGTCCGATCCTGTTT 3796
      |||:::|||||
1018 .....LeuasnAspGlyIn..... 1022
3797 CGCACAAACCGACCGGAAACCTTCAGACGCGATCGCACTCGGCA 3846
      ::::: ||| ||| ::::: |||:::
1023 ..AsnLysThrThrThrasnGlnpheIleasnGlnLeuGlyGlyPle 1038
      |||::: |||::: |||::: |||::: |||::: |||:::
3847 CGGCTTGCCGACGCGTCCGTTTCGGCAATACGCGATCGGATTCGA 3896
      |||::: |||::: |||::: |||::: |||::: |||:::
1039 TyrIysPheHisAlaGlu.....GlnLeuGlyAspPheTh 1050
3897 CATCGGATATCGCGCGCGCGCGTTCAGTACGCGACCTTTCACACG 3946
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1050 rLeuGlyIleMetGlyTyrAlaAsnAlaLysGlyLysThrIleAsn 1067
3947 GCATCAGAGGCAAAATCCGCGCGCGCGTTCAGTACGCGATTCAGCA 3996
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1067 YrThrSerAsnLysAlaAlaLysAsnThrLeuAsp..... 1078
3997 AGATACCGCGAGTTCGCGGATTCGCGATCGACGACATCGCGCGC 4046
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1079 GlyTyrSerValGlyValTyrGlyThrTyrPyrGlnAsnGlyGluAsnAl 1095
4047 AACGCGCTATTCGTCACAAAA..GCGGATACCGATACGAAACGCA 4093
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1095 aThrGlyLeuPheAlaGluThrTyrMetGlnTyrAsnThrPheAsnAlas 1112
4094 ATATCGCACCCCGCGCTTCGATTCACACGCGTACCGC..... 4131
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1112 erValLysGlyAspGlyLeuGluGluGlyTyrAsnLeuAsnGlyLeu 1128
4131 ..... 4131
1129 ThrAlaSerAlaGlyGlyTyrAsnLeuAsnValHisThrTyrPhe 1145
4132 ....GCGGCAATTAGCGAGATTATTCATTCACACG.....GCGC 4168
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1145 rProGluGlyIleThrGlyGluPheThrPleuGlnProHisLeuGlnAlay 1162
4169 AACACATTCACACGCGCTTATTCAGCCTGTCTATACGAGTCCGCT 4218
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1162 aLTrpMetGlyValThrProAspThrHis..... 1171
4219 TCCGCGCAAGTCCGAACGCGCTCAATACCGCGTATTCGCGCAGATT 4268
      ::::: |||::: |||::: |||::: |||::: |||::: |||:::
1172 .....GlnGluAspAsnGlyThrValGlnGlyAl 1182
4269 CGGCAAAACCCG.....AGTGGCAATGGGCGC 4297
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1182 aGlyLysAsnAsnIleGlnThrLysAlaGlyLysLeuArgAlaSerTyrLys 1199
4298 TAAACGCCCAATCAAAAGTTTCACGCTCCCTCCACGCTCCGCGCGC 4347
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1199 aLlySer.....ThrLeuAspLysAspThrLysArg 1210
4348 AAGGGCGCGCAATTCGACGCGAG.....CACAGCGCGG 4382
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1211 pheArgProTyrIleGlnAlaAsnTrpIleHisAsnThrHisGlnPhe 1227
4383 CATCAATTA 4392
      |||::: |||::: |||::: |||::: |||::: |||::: |||:::
1227 yValLysMet 1230

```

seq\_name: SwissProt\_40:BIGA\_SALTY

seq\_documentation\_block:

ID BIGA\_SALTY STANDARD; PRT; 1953 AA.  
AC P25927; P25928; O9XC03;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

DE putative surface-exposed virulence protein bigA precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subphylum; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-14028.
RA Stojiljkovic I., Valentine P., Heffron F.
RT "Salmonella typhimurium rns homolog."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
[3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=91100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cygB plasmid to overcome limiting srl gene
RT cofactor."
RL J. Bacteriol. 173:325-333(1991).
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 414 and 732.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF133696; AAD39458.1; -.
CC EMBL; AE008859; AAL22340.1; -.
CC EMBL; M64606; AAA27042.1; ALT_FRAME.
CC EMBL; M64606; AAA27043.1; ALT_FRAME.
CC PIR; C39200; C39200.
CC PIR; D39200; D39200.
CC
CC StyGene; SG10437; bigA.
DR Virulence; Repeat; Signal; Complete proteome.
KW SIGNAL
FT 1 27
FT CHAIN
FT 28 1953
FT
FT DOMAIN
FT 101 252
FT REPEAT 101 103
FT REPEAT 104 113
FT REPEAT 104 122
FT REPEAT 114 122
FT REPEAT 123 133
FT REPEAT 134 144
FT REPEAT 145 155
FT REPEAT 156 166
FT REPEAT 167 177
FT REPEAT 167 188
FT REPEAT 178 188
FT REPEAT 189 199
FT REPEAT 200 210
FT REPEAT 211 221
FT REPEAT 222 232
FT REPEAT 233 243
FT REPEAT 244 252
FT REPEAT 207 207
FT CONFLICT
FT 514 514

```

15 X 11 AA TANDEM REPEATS.  
1 (INCOMPLETE).  
2 (INCOMPLETE).  
3 (INCOMPLETE).  
4.  
5.  
6.  
7.  
8.  
9.  
10.  
11.  
12.  
13.  
14.  
15 (INCOMPLETE).  
D -> DRGDDVTPPD (IN REF. 1).  
A -> R (IN REF. 3).

FT CONFLICT 1698 1698 D -> N (IN REF. 1).  
 FT CONFLICT 1795 1798 OXLE -> TLO (IN REF. 1).  
 FT CONFLICT 1836 1837 SA -> T (IN REF. 1).  
 SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

## alignment\_scores:

Quality: 245.00 Length: 1687  
 Ratio: 0.314 Gaps: 80  
 Percent Similarity: 46.236 Percent Identity: 19.265

## alignment\_block:

US-09-303-518D-653 x BIGA\_SALTY ..

Align seg 1/1 to: BIGA\_SALTY from: 1 to: 1953

```

217 GTTTACACAAAAAGGGAGTGGTCGCAATCCATGACGAAAGCCC 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 ValAspAsnAlaThrGlyAlaLeuIleSerGlyAsnGlyThrThrAs 429
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
267 GATGATGATTTTCTGTCGTCATCGCTACGCGTGGCGCATTTGGCG 315
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
429 nPhaAlaGlyAspIleAlaValSerGlyGlyGlyThrAlaIleIleA 446
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
316 .GGCGATCAATATTATTGTCAGCGTGCACATACGCGCTTAAACAT 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
446 spGlyAspAsnAlaThrIleLeys.....AsnThrGlyThrSerAsp 459
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
364 GTTGAT...TTGTGCGGAGGAGGACCAATCCCGATCAGCAGCGCTTTC 410
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
460 IleSerGlyAlaGlySerThrGlyThr..... 468
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
411 TTACCAATTGTGAAAAAATATTATATAA..... 441
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
469 .....ValIleAspGlyAsnAsnAlaArgValAsnAsnAspGlyAspM 483
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
442 .....GCAGGAGCTAACGGCCATCCTATGCGCGCTTAT 477
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
483 ethrIleThrAspGlyGlyThrGlyGlyIleThrGlyAspAsnVal 499
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
478 CATATGCGCGTTTGACAAATTTGTCACAGATGCAGAACCTGTGAGAT 527
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
500 ValIleAspAsnAlaGlySer.....ThrIleValSerGlyAlaAspAl 514
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
528 GACCACT...TATATGATGCGGTGCAATACCTGATTTAAATTAATACC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514 athrAlaLeuTyrlleGlyIleGlyIleAspAsnAlaLeuValIleAsnGlyGlyA 531
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
575 CTGATGCTGTGCAATCGAGCA...GGCAGACAAATATTGGCGGCTGTAT 621
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
531 snGlnThrIleSerGlyGlyAlaValGlyThrArgIleAspGlyAspAsp 547
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
622 GAAGACGAACCAATACCGCGCAAGTTCATATTCATATTGCAAGCGCAT 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
548 AlaIleThrThrAsnThrGlyAspIleAlaValAspGlyAlaGlySerAl 564
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
672 TTCCTGGCTCGTGGGCGCAATACC..... 696
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
564 alaValIleIleAsnGlyAspAsnGlySerLeuThrGlnAlaGlyAspL 581
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
697 .....TTGCAACAAATGATGATCAGCTGT 720
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
581 eulLeuValThrAspGlyAlaMetGlyIleIleThrTyrlleGlyThrGlyAsn 597
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
721 GGCAGACGTAAGTATAGTACGCAAAAAATTAACATAC...CCATATGG 767
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 GluAlaValAsnThrGlyAsnAlaThrValArgAspAlaAspSerValGI 614
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
768 TTTTTCACACAGCA.....GGCTCATTT.....GGCGC.... 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 yPheValAlaGlyIleGlyAsnThrPheLeuAsnGlySerAlaIleA 631
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
799 .....AGTGGCTACCAATGTTTATCTATGATGCCCAAAAGCA 837
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

631 spValSerLeuAsnGlyThrGlyAlaLeuValSerGlyAspMetSerGln 647
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
838 AACTGGTTA.....ATTAAT..... 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
648 ValThrLeuAspGlyAspIleAsnValAlaSerValGlnAspSerGluGI 664
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
855 GGTATTGCAA.....ACAGCAACCCCTATATAGCAAAAAACCAATGGCT 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 yAlaPheSerSerAlaThrGlyValSerValSerGlyAspSerAlaVal 681
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
899 TCACGCTA.....GTTCTAAAGATTGGTTCTATATATGA 933
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
681 AlaSpIleThrGlyAsnValAsnIleSerAlaAspTyrlleGlnAspAsp 697
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
934 ATCTTTGCTGAGATACCATTCAGATTTCTACGAAACCAATCAAAATGG 983
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
698 LeuAlaAlaGly.....AlaProPoleuThrGI 707
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 GAAATACCTTTTAAACACAAATAT.....ATGCGCAGGAA 1021
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
707 yValValValGlyGlyAsnGlyAsnThrValThrLeuAsnGlyAlaLeuA 724
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1022 AATGATGCGCAACATAACACACTTCTCTACCTTATAGATTAANAACA 1071
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 snIleAspAsnAspLeuSer.....AlaThr 733
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1072 CGAACCGTTCATTTGTTAATGTTTCTTATCCGACAGCAAGAACAC 1121
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
734 GlyGlyGlnTyrlleuAspValValGlyLeuSerValThrGlyAspAspAs 750
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122 TGTTATATCTGTCAGCTGGGCTCAC.....AGTATGACCCA 1162
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
750 nAspValGluIleAspGlyGlyIleAsnIleThrHisSerGluAspProl 767
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1163 GACTCAATATAGAGAAATAATTCCTTATTTGAC...AAGGAAAGGT 1209
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767 euAspGlyThrSerAlaAspIleThrGlyIleSerValSerGlyAsnSer 783
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1210 GAATTGATCTT.....ACGACCAATCAACCAAGCGCGG 1247
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 ThrValThrLeuAsnGlyHisSerThrIleAspThrAsnThrValAlGI 800
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1248 CGCTTTGATTTGAGGTATTTTACGGTCTGCCCTAAAAACACAGCAA 1297
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
800 Y.....GlyHisValValLeuAlaArgValAsnAsn.... 810
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1298 CGTGCAAGCGCGCGGCTTCATATCAATGATGCAATGCTGTTACTTGG 1347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
811 .....GlyGlySerLeuIleGluGlyAspAspSerValAlaAspVal 824
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1348 AAAGTAAACGGCGTGGCA.....AACGACCGCTGTC 1379
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825 AsnValSerTyrlleProThrGlyTyrlleThrTyrlleAsnAlaLeuIleuMe 841
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1380 CAAATGCGCAAGACGACGCTGCTGTTCAAGCAAGGAGGA..... 1422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
841 tAlaAspGlyGluGlyThr...SerIleGluAsnGlyGlyAspIleHis 857
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1423 .....AACCAAGGCTCGTCAAGCTG 1443
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
857 erHisGlyValTyrlleSerValIleArgAlaAspAsnGlySerGluValSer 873
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1444 GGCAGCGTAAAGTCATCTAGATACACAGCGGACGATCAAGCAAAAA 1493
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 AsnSerGlyAspIleLeuValTyrlleThrThrSerThrAsnSerSerGluAs 890
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1494 ACAAGCTTTAGTGAATCGGCTTGTGTCAGCGGACGAGGCGCTGCAAC 1543
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 pArgAla.....AlaIleThrArgAlaSerGlyGluGlySerAlaValH 905
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1544 TGAAT.....GCGGATATCAGTCAACCCGCAACAAACTCTATTTC 1584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





```

3112 TGGCGTTATCAGCTTATCCGCAAGACGGAGTCCGCTG..... 3153
      :::: ||| :::: ||| :::: ||| :::: |||
1451 GilyThrThraAlaIleIleAsnnglyIleAsnLeuLeuCysAs 1467
3154 .....C 3154
1467 pSerGlyCysAspIleTyrAlaProGlyThrThrGlyThrGlnAspH 1484
3155 ATAAATCCGGTCAAGACAGAGCTTCCGCAACCTCGGACAGCGGGA 3204
      |||| :::: ||| :::: ||| :::: |||
1484 lAsnnglyThrAlaAspIleValIleProAsp.....AlaThrThra 1498
3205 GAACAGAGCGCGCTTGACGGCAAAACAGCAACTGCGCCCAACA 3254
      |||| :::: ||| :::: ||| :::: |||
1499 ProThrGlnGlySerIleProThrProProAlaAspProAlaProG 1515
3255 AAGGCGGAAAA..... 3267
      ||| :::: ||| :::: ||| :::: |||
1515 nGlnLeuSerAsnTyrIleValGlyThrAsnAlaAspGlySerSerG 1532
3268 .....GACACGGCGCAAGCCTTGAC 3288
1532 hrLeuLysAlaAsnAsnLeuValIleGlyAspAsnValLysValAs 1548
3289 GCGCTGATTGCGCGCGCGCAATGCCGAAAGCGAAGAGTGTTC 3338
      :::: ||| :::: ||| :::: ||| :::: |||
1549 GlyPheThrSerGlyThrAlaAspThrThraValValAlaAspAsn 1565
3339 GCAACCGCGCGCGAGCGGCGGGAAT.....GCCGCAATTA 3379
      :::: ||| :::: ||| :::: ||| :::: |||
1565 eThrGlySerAsnIleGlnGlyAlaAspAsnIleThrSerThrSer 1582
3380 TCCAGCGGAGAGAGAAAAACGGGTGACAGCGGATAAAGACCGCC 3429
      :::: ||| :::: ||| :::: ||| :::: |||
1582 alTrrAsnAlaGlnGlySerGlnAspAlaAspGlyAsnValAspAla 1598
3430 TTGGCGAAACAGCGCAAGCGGAAACCGCGCGCTACCGCGCTTCC 3479
      :::: ||| :::: ||| :::: ||| :::: |||
1599 MetThrLysAsnAlaTyrAlaAspValAlaThrAspSerSerVal 1613
3480 CCGCGCGCGCGCGCGCGCGGATTGCGCAACCGCGCCCAACCGC 3529
      ||| :::: ||| :::: ||| :::: |||
1614 .....SerAspValAlaGlnAla..... 1619
3530 AACCCCAACCGCAGCGCGATGATCGCTTATGCCATAGCGGTTTG 3579
      ||| :::: ||| :::: ||| :::: |||
1620 .....LeuAspAlaGlyTyrThrAsnAsnLeu 1629
3580 .....AGTAATTTCCGCCAGCGTCAA 3602
      :::: ||| :::: ||| :::: ||| :::: |||
1630 TyrThrSerLeuAsnValGlyThrThraGlnLeuAsnSerAlaLeu 1646
3603 CAGCGTTTGGCGGTACAGAGCAATGGACCGGTTTCCCAAGACC 3652
      :::: ||| :::: ||| :::: ||| :::: |||
1646 GlnValSerGlyAlaGln.....AlaThrThraValPheArgIleAla 1661
3653 GC..... 3654
      ||| :::: ||| :::: ||| :::: |||
1661 rgValLeuSerAsnArgPheThrMetLeuAlaAspAlaAlaProGln 1677
3655 .....CGCAACCGCGTTGGCAACCGCGATCCG..... 3684
1678 LysAspGlyLeuAlaPheAsnValValAlaLysGlyAspProArgAla 1694
3685 .....GACACCAACACTACCGTTCCAGAGATTCCGCGCTAAC 3724
      |||| :::: ||| :::: ||| :::: |||
1694 uLeuGlyAsnAspThrGlnTyr.....AspMetLeuAlaLeu 1707
3725 GCCAACAACCGACCTGCG.....CAATCGGT 3753
      |||| :::: ||| :::: ||| :::: |||
1707 rgGlnThrLeuAspLeuThrAlaSerGlnAsnLeuThrLeuGlyTyr 1723
3754 ATGCAGAAACCTCGGACGCGCGCTCGCATCTCTGTTTGCACAA 3803

```

```

      :::: ||| :::: ||| :::: ||| :::: |||
1724 lIleAlaTyrLeuAspGlyAspGly.....SerLysTh 1734
3804 CCGGACCGGAAACACCTTCGACGACGATCGCAACTCGGACGCGCTG 3853
      ||| :::: ||| :::: ||| :::: |||
1734 lAlaGlyAspAsnGlyLeuThrGlyGlyTyrSerGlnPhePheGlyLeu 1751
3854 CCCACGGTCCGCTTTCGGCAATACGCAATCGGC..... 3888
      |||| :::: ||| :::: ||| :::: |||
1751 yshISerMetAlaPheAspGlu.....GlyLeuAlaTrrAsnAsnSer 1766
3889 AGTTTCGACATC...GGCATCAGCGCGCGCGGCTTTAGTAGCGGAC 3935
      |||| :::: ||| :::: ||| :::: |||
1767 ArgTyrAspValHisAsnLeuAspSerSerArgSerValAlaTyrGly 1783
3936 CCTTCAGACGGCATCAGAGGCAAAATCCGCGCGCGCTGCATATAC. 3984
      :::: ||| :::: ||| :::: ||| :::: |||
1783 pValAsnLysIleAlaAspSerAspMetArgGlnTyrLeuGlnPhe 1800
3985 .....GGCATTCAGCAAGATACCGCGAGTTTCGGCGATTCGGC 4026
      ||| :::: ||| :::: ||| :::: |||
1800 rgSerGlnGlyAlaLysThrPheThrMetMetGly...AspAlaLeu 1815
4027 ATGCAGCGCATCGCGCAACGCGCTATTGTCGCAAAAGCGGATTA 4076
      :::: ||| :::: ||| :::: ||| :::: |||
1816 ValThrProTyrAlaGlyValLys.....PheArgHisThrMetGlu 1830
4077 CCGATACGAAACGTCATATCGCCACCGCGGCTTGCATTCACCGCT 4126
      |||| :::: ||| :::: ||| :::: |||
1830 pGlyTyrLysGlnArgSerAlaGlyAspPheAsnLeuSerMetLys 1847
4127 ACCGC.....GCCGCAATTAAGCAGATTTATCA 4155
      :::: ||| :::: ||| :::: ||| :::: |||
1847 LysGlnGlyThrAlaValAspSerIleValGlyLeuLysLeuAsp 1863
4156 TTCAAACCGCGCAACACATTTCCATCACG.....CCTATTT 4193
      ||| :::: ||| :::: ||| :::: |||
1864 GlyLysAspGlyTrrSerAlaThrAlaThrLeuGlnGlyLysPro 1880
4194 GAGCTTCCTATACGATGCCGCTTCGCGAAGTCCGACGCGGCTCA 4243
      |||| :::: ||| :::: ||| :::: |||
1880 uSerTyrSerLysSerGln.....ArgThrAlaSerL 1891
4244 ATACCGCGGTATGGCGCAGGATTTGCGCAAAACCGGATGGGAA 4293
      :::: ||| :::: ||| :::: ||| :::: |||
1891 euGlnGlyAlaAlaGlyGlnSerPheGlyValAspAspGlyGlnGly 1907
4294 GCGGTAAACCGCGAAATCAAGGTTTCACGCTGCCCTCCACGCTCC 4343
      ||| :::: ||| :::: ||| :::: |||
1908 Gly.....GlyValAsnGlyLeuAl 1914
4344 CGCCAGGCGCGCAATGGACGCGACGACGCGGCGGATTAATTAG 4393
      :::: ||| :::: ||| :::: ||| :::: |||
1914 aThrIleGlyValLysTyrSerSerAsnAspThrAlaLeuHisLeu 1931
4394 GCTACCGCTGG 4404
      ::||| :::: ||| :::: |||
1931 latYrGlnTrr 1934

```

seq\_name: SwissProt\_40:YEJO\_ECOLI

```

seq_documentation_block:
ID YEJO_ECOLI STANDARD: PRT: 863 AA.
AC P33924: P76450: Q47291:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein yefJ.
GN YEJO OR B2190.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

```



```
2371 ACGGCAAGCCAC.....CAAAAGCAACCT 2399
      |||||
314 GlnAlaAsnIValItyrAspLeuGlyArgSerTyrGlnAsnIleIleI 330
      |||||
2400 CAGCCTCGTGGCAATGCCCAACCAATTTAATCAAGCCACATTAACG 2449
      |||||
330 eTrhTyrSerSerLysAla...IleSerGluAsnMetValIleAsnAsnG 346
      |||||
2450 GCAACACATCGGCTCGACATCGCTTCAATTTAATCAAGCAACACGCC 2499
      |||||
346 LyrArgAlaAsnValITrpAlaGlyIThrMetValAsnValSerValArgGly 362
      |||||
2500 GTACAAAGCGGCTGTACGCTTCCGACACGCTAAGGCAACGTAAG 2549
      |||||
363 .....AsnAspGlyIleLeuGluValMetLysProGlnIleAsnTyrAl 377
      |||||
2550 CCATTCGGCATCAAGGCAATGCTTCCTCCAGCCGTAAGCGCATTTCC 2599
      |||||
377 aProAlaMetLeuValIglYlyValValSerGluGlyAlaSerPhe. 393
      |||||
2600 ATTTGAAAACAGCGCTTACCGGAAATACGCGGCAAGATAG 2649
      |||||
394 .....ArgThrHisGlyAlaValAspThr 401
      |||||
2650 GCA.....TTACACTTAAAGACAGCAATGAGCGCTGCGTGGG 2690
      |||||
402 SerLysAlaAspValSerLeuGluAsnSerValITrpThrIleIleAla 418
      |||||
2691 C.....ACGGAATTAGCAATTTTAAC 2713
      |||||
418 pIleIThrIThrAsnGlnAsnIThrLeuLeuAsnLeuAlaAsnLeuAla 435
      |||||
2714 TTGACACAGCCACCAATTAACACTCAATTCGCTATGACAGCATGGCG 2763
      |||||
435 eSerAspAlaAsnValIleMet..... 442
      |||||
2764 GGGGCGCAACCGGCGAGTGGCGAGATGGCGCGCGCGCTGCGCGCG 2813
      |||||
443 .....MetAspGluProValIThrArgSer..... 450
      |||||
2814 TTCCCTATTATCCGTTACGCGGCAACTTGGCGAGAAATCCGTTCAACA 2863
      |||||
451 .....SerValIThr.....AlaSerAlaGluAsn...PheIleT 461
      |||||
2864 CGCTGACGTTAAAGGCAATTGAACGCTCAGAGCAATTCGCTTTATG 2913
      |||||
461 hTrLeuIThrIThrAsn...ThrLeuSerGlyAsnGlyAsnIThrGlyGlnAl 476
      |||||
2914 TCGGAACCTGTGGCTACCGCGGCAACATGAAGCTGGCGGAAGTTC 2963
      |||||
477 ThrAspMetAlaAsnHisGlnSerAspGlnLeuAsnValIThrGlyGlnAl 493
      |||||
2964 CGAAGCACTTACACCTTGGCTGCAACATACCGGCAACGCAACCGTAA 3013
      |||||
493 aThrGlyAspPheLysIlePheValIThrAspIThrGlyAlaSerProAla 510
      |||||
3014 GCTCTGAGCAATTTGACGGTAGTGAAGAAAGACACACACCGCTGTCC 3063
      |||||
510 lAgIlyAspSerLeuIThrLeuVal.....ThrIThrGlyGly 521
      |||||
3064 GAAATCTTAAATTTCACCTCGCAAAACGACAC.....GTGATCGCGG 3107
      |||||
522 GlYAspAlaAlaPheIThrLeuGlyAsnAlaGlyValValAspIleGln 538
      |||||
3108 CGCATGCGCTTATCAGCTTATCCGCAAGAGCGGAGTTCCGCTGCATA 3157
      |||||
538 yThrITyrGluITyrIThrLeu..... 545
      |||||
3158 ATCCGCTAAAGACAAAGACTTTTCGAAACTCGGAGCGGAGAA 3207
      |||||
545 ..... 545
3208 ACAGAGCGCGCTTGACGCAAAACAGCACACTTCCCGCCAAACACA 3257
      |||||

545 ..... 545
3258 GCGGGAAGAAAGACAGCGGCAAAAGCCTTGACGCGCTGATTCGGCGGCG 3307
      |||||
546 .....AspAsn.....GlyA 549
      |||||
3308 GCATGCGCACGGAAGAAAGCAGAAAGTGTTCGCGAACGCGCGGAGGCA 3357
      |||||
549 snHisSerITrpSerIThrLeuAlaGluAsn..... 557
      |||||
3358 GCGCGGGAAGAAATCGCGGATATGACAGCGGAGAGAAAGAAAGCGGT 3407
      |||||
557 ..... 557
      |||||
3408 GCAGGCGGATAAAGACACCGCTTGGCGAAACAGCGGCAACGCAACCC 3457
      |||||
558 .....ArgAlaGlnIleT 562
      |||||
3458 GCGCGGCTACGACCGCTTCCCGCGCGCGCGCGCGCGGATTTG 3507
      |||||
562 hTrProSerIThrIThr..... 566
      |||||
3508 CCGCAACCGGACGCCCAACCGCAACCCCAACCGGACGCGACCTGATCAG 3557
      |||||
567 .....AspValIleuAs 570
      |||||
3558 CCGTTATGCCAATAGCGGTTTGAAGTATTTCCGCGCACGCTCAACAGCG 3607
      |||||
570 mMetAlaAlaAlaGlnProLeu...ValPheAspAlaGluLeuAspIThr 585
      |||||
3608 TTTTCGCGCTACAGGCAATTTGACCGCGCTGTTGCCGAAGACCGCGCG 3657
      |||||
586 .....ValArgGluIThrGlyLeuGlySerValIlyGlyValSerTyrAsp 599
      |||||
3658 AACGCGCTTTGGACAAAGCGGATCGCGGACACAAACACTACCTGTCGA 3707
      |||||
600 ThrAlaMetITrpSerSerAlaIle...AsnIThrArgAsnAsnValIThrITh 615
      |||||
3708 AGATTTCGCGCGCTTACCGCCCAACCAAC..... 3735
      |||||
615 rAspAlaGlyValaGlyPheGluGlnIThrIThrGlyLeuIThrLeuGlyI 632
      |||||
3736 ..GACCTGCGGCAAAATCGGTTATGCAAAACCTCGGCGAGCGGCGGCTC 3783
      |||||
632 lAspSerIThrArg.....PheSerITrgIuGluSerIThrIThrArg 645
      |||||
3784 GGCATCTGTTTTCGACAAACCGGACGCGAAACACTTTCGACGACGAGCAT 3833
      |||||
646 GlyLeuIlePheGlyITyrSerHisSerIThrIleGlyPheAspITrgIy.. 661
      |||||
3834 CGGCACTCGGACGCGCTTGGCCACGCTGCCGTTTGGGCAATACGCGCA 3883
      |||||
662 .....GlyLysGlyAsnI 666
      |||||
3884 TCGGAGGTTTCGACATCGGATGACGCGGCGCGGCTTTAGAGCGGC 3933
      |||||
666 lAspSerITyrIThrLeuGlyAlaITyrAlaGlyITrgIuHisGlnAsnGly 682
      |||||
3934 AGCCTTTCAGACGCG.....ATCAG 3953
      |||||
683 AlatyITyrAlaITrgIyValITryValITryAspITrgPheAlaAsnIThrIle 699
      |||||
3954 AGGCAAAATCGCGCGCGCTGCTGCATTTACGCG..... 3987
      |||||
699 sGlyLysMetSerITrgIyAlaIThrAlaPheGlyAspITrgITrgAsnSerITrg 716
      |||||
3988 .....ATTCAGGCAAGATACGCGGACGCTTTCGCGGAGTTC 4023
      |||||
716 LyAlaGlyAlaHisValaGluSerITrgITrgPheITrgITrgITrgITrgITrg 732
      |||||
4024 GGCATCGAACCGGACATCGGCGGCAACGCGCTATTTCGTCGCAAAACGCGA 4073
      |||||
```



481 ATGCCGCTTTGCACAAATTTGTCCACAGATGCACCAACCTGTTGAGATGAC 530  
470 ThrValAspAsnLysGlyThrMetThrValThrAspProGlnSerIleGly 486  
531 CAGTTATATGATGGTGGAAATACGCTGATTTAAATTAATACCTGATC 580  
486 ylleGlnIleAspGly..... 491  
581 GTGTTCAATTCGAGCAGCAGACATATTTGGCGTCTGTGTAAGCA 630  
492 .....AspGlnAlaIle 495  
631 CCCAATACCGCGCAAGTTCATATCATATTCAGACGCAATATTCCTGGCT 680  
496 ValAsnAsnGlnGlyGlnSerThrIleThrAsnGlyGlyThrGlyThrGly 512  
681 CGTGGTGGCAATACCTTTCACAAAATGATCAGGTGGTGACAGATC. 729  
512 ntleAsnGlyAsnAspAlaThrAlaAsnSerGlyLysThrThrValA 529  
730 .....AACTAGGTACGAA 744  
529 spGlyLysAspSerThrGlyThrLysIleAlaGlyAsnIleGlyIleVal 545  
745 AAAATTAAACATAGCCCATATGGTTTATTCACCAACAGAGAGC..... 786  
546 AsnLeuAspGlySer.....LeuThrValThrGlyGlyAlaHisGly 559  
787 .....TCATTGGCGCAGCTGCTCA.....CCAATGT 814  
559 yValGlnAsnIleGlyAspAsnGlyThrValAlaAsnLysGlyAspIleVal 576  
815 TTATCTATGATGCCCAAAAGCAAAAGTGTATTAATGAGG..... 855  
576 alValSerAspThrGlySerIleGlyValLeuIleAsnGlyGlyIleValA 592  
856 ..GTTTGCAAACAGCAGCAACCCCTATATA..... 882  
593 ThrValSerAsnThrGlyAspValAlaAsnValSerAsnGlyAlaThrGlyPh 609  
882 ..... 882  
609 eSerIleThrThrAsnSerGlyLysValSerLeuAlaGlySerMetGlnV 626  
883 .....GGAAGCAATGGCTGC 900  
626 alGlyAspSerThrGlyValAspLeuAsnGlyAsnAsnAsnSerVal 642  
901 CAGCTAGTTCGTAAGATGGTTCTATGATGAATC..... 936  
643 ThrLeuAlaIleLysAspLeu.....LysValValGlyGlnLysAl 656  
937 .....TTGCTGAGATACCATTCAGTA..... 960  
656 aThrGlyIleAsnValSerGlyAspAlaAsnThrValAsnIleThrGlyA 673  
961 .....TTC 963  
673 snValLeuValAspLysAspLysThrAlaAspAlaIleGlyLysrhe 689  
964 TACGAACCCATCAAAATGGAAATCTTTTAAACGACATATTAATG 1013  
690 pheAspProSerValGlyIleAsnValIleGlySerAspAsnAsn..... 704  
1014 CGCAGGAAATATGATGCCAAACATAAACATATCTCTACCTTATGAT 1063  
705 .....ValThrLeuAspGlyLysLeuThrValValSer...AspSerGlnV 719  
1064 TAAAAACAGAACCTTCAATTTGTTAAATGTTTCTTATCCGAGACACA 1113  
719 alThrSerValGlnSerAsnLeuPheAspGlySerAlaGlnLysThrSer 735  
  
1114 AGAACCTGTTTATCATCTCGAGGTGGGTCAC..... 1149  
736 GlyLeuValValIleGlyAspGlyAsnThrValAlaMetAlaGlyIle 752  
1150 .....AGTTATCGACCCAGACTGAATATGAGAAATATTTCT 1189  
752 uGlnLeuIleGlyGlnLysAlaAlaLeuAlaAspGlySerGlnValHis 769  
1190 TTATTGACAAAGAAAGT.....GAATGATACTTACCAACAATC 1233  
769 eLxLeuArgThrGlyTyrSerTyrThrSerValIleValValSer..... 783  
1234 AACCAAGCGCGCGCGTGTGTTGAGGTAAATTTACGCTGCC 1283  
784 .....GlyGlnSerSerValIleLeuAsnGlyAspThrIleSerGly 798  
1284 TAAAAACAGAAACGTGCAAGC.....GGCG 1312  
798 yGlnPheProLeuGlyPheAlaGlyValIleArgValGlnAspLysAlaL 815  
1313 GCGTTCATATTCAGTATGACAGTACCTGTTACTTGAAGTAAACGCGTG 1362  
815 eLxLeuGlnIleGlySerGlyAlaThrLeuThrMetGln..... 827  
1363 GCAACGACCGCGCTGCCAAATCGCAAGGACAGCTGCTGTTCAA.. 1410  
828 AspIleAspSerPheGlnHisGlyThrArgThrValGlnIleGlnAs 844  
1411 .....GCCAAAGGGGAAACCAACGAGCTGGCGAGCGGG 1444  
844 nLeuGlyPheAlaPheValThrGlyGlnAsn.....ThrThrGlyIleA 859  
1445 GCGACGCTTAAGTCACTTACATCAGCGCGAGATCAAGGCAAAAA 1494  
859 snSerGlyThrIleSerLeuGln.....AsnGlyLysAsp 871  
1495 CAAGCCTTAGTAATC..... 1512  
872 ProAlaProSerProIleValLeuLeuAlaThrAsnGlyLysSerAlaTh 888  
1513 .....GGCTTGTACAGCGGAGGAGCGGTGCACACTGAATCCGATA 1555  
888 rAsnAlaGlyThrIleThrGlyLysValThrGlnGlnHisSerValPheA 905  
1556 ATCAGTTCAACCC...GACAACTCTATTGCGCTTCCGCGGAGCT 1602  
905 snLysTyrSerThrGlyThrSerAsnSerPheIlePheAsnAsnAspVal 921  
1603 TTGATTTGAACGG...CATTCCTTTCGTTCCACGCAATTCAAATAC 1649  
922 SerSerIleThrGlyLeuValAlaGlnSerAsnSerThrIleIleAsnTh 938  
1650 CGATGAACGGCGATG..... 1665  
938 rAspSerGlyIleIleAspLeuTyrGlyArgGlySerValGlyMetLeuA 955  
1666 ..ATTGTCAACCAATCAACAGACAAAGATCCACCTTACATT..... 1707  
955 lAlaIleAspSerThrAlaGlnAsnGlnGlySerIleThrIleAspSer 971  
1708 .....ACAGCAATTAAGATTTACTACA.. 1731  
972 MetThrValAspAlaAsnAspThrThrAlaMetArgAspIleAlaSerAs 988  
1732 .....ACGGCAATTAACAACACTTGATAGC.... 1758  
988 nSerAlaIleAspPheGlyThrGlyValGlyValGlyThrAspSerTys 1005  
1759 .....AAAAAGAAATGGCTACAC..... 1779  
1005 eArgAlaGlyLysAsnAlaThrAlaIleAsnGlnLeuGlyGlyValIle 1021  
1780 .....GTTGGTTGGCGAGAAAGATGC 1802

```
1022 ThrIleTyrAsnAlaGlyAlaGlyMetAlaAlaIleTyrGlyAlaSerAsnThr 1038
      :::: :::::
1803 AACCAAAACGACGCGCGCTCAATCTG.....AATTACCAAC 1840
      :::: :::::
1038 rValIleAsnGlnGlyThrIleAsnLeuGluLysAsnGlyAsnTyrAsp 1055
      :::: :::::
1841 CGAAGAAGAGGATCGCACTTA.....CTGCTTCGGGGGA 1878
      :::: :::::
1055 sPserLeuAlaIleAsnThrLeuValGlyMetAlaValTyrGluHisGly 1071
      :::: :::::
1879 ACAATTAAACGCG.....AATTACGCAAAACACG 1913
      :::: :::::
1072 ThrAlaIleAsnAspGlnThrGlyValIleAsnIleAsnValGlyThrGln 1088
      :::: :::::
1914 CAACCTGTTTTTCAGC...GGCAACGCGACCGCGCTACATCATTT 1960
      :::: :::::
1088 yGlnAlaPheTyrAsnAspGlyThrGlyThrIleValAsnTyrGlyThrI 1105
      :::: :::::
1961 TA.....GGACGCGGGTGTCAAAATGGAA 1986
      :::: :::::
1105 lEcysThrPheGlyValGlySerGlnSerGlyAsnGluTyrAsnAsnThrAsp 1121
      :::: :::::
1987 GGTATCCCA.....CAAG 2000
      :::: :::::
1122 AspPheThrSerLeuIleTyrThrGlyGlyAspThrIleThrArgSerGln 1138
      :::: :::::
2001 AGAATCTGTGCGACACGATGATGACCGCACATTAA..... 2043
      :::: :::::
1138 yGlnThrValThrLeuAsnLysSerAlaIleValThrAspLysLeuAlaG 1155
      :::: :::::
2044 .....GCGGAAACTTCATATTCAGCGGACAGCGGTGTTTC 2085
      :::: :::::
1155 lYAsnValValAsnSerGlyThrLeuSerGlyAspGlnIleThrValSer 1171
      :::: :::::
2086 CGCAATGTTGCCAAAGTGAGCGCATTTGCAATTAACATCACGCCCA 2135
      :::: :::::
1172 SerGlyLeuLeuGlnAsnThrSerGlyGlyIleIleAsnAsnLeuValLy 1188
      :::: :::::
2136 ACCAGTTTCCGGTTCGCGCGCATCAAGCACACATCTGTACAGCTT 2185
      :::: :::::
1188 sLeuAspLysGlyAlaValIleLysAsnAlaGlyValMetThrAsnAsnV 1205
      :::: :::::
2186 CGGACTGGAAGGT.....CTGACAGTTGTACGAAAAACCATTAAC 2229
      :::: :::::
1205 aLAspValSerGlyGlyIleLeuAsnAlaGlyGluMetThr..... 1219
      :::: :::::
2230 GACGATAAGTGATGCTTCATTGAGCAAGCCGACATCAGAGCAATGT 2279
      :::: :::::
1220 .....AlaGlnIleThrMetAsnAl 1226
      :::: :::::
2280 GACGCTTGCGATCAAGCTCATTTAAATCTCACAGACTTGCACACATCA 2329
      :::: :::::
1226 aGly...AlaAspSerSerLeuValAsnAsnThrGly.....ThrIleA 1240
      :::: :::::
2330 AC.....GCGCAT..... 2337
      :::: :::::
1240 sNlysIleValGlnAsnAlaGlyValPheAsnAsnSerGlySerValThr 1256
      :::: :::::
2338 .....CTTAGTCACAGCGGA.....GACAGCCTATATAC 2366
      :::: :::::
1257 GlyArgMetMetSerAlaGlyValPheAsnAsnGlnThrAspGlyVal 1273
      :::: :::::
2367 GGTTCAGCGCAACGCGCAACCGCAACCTCGTGGGCAATG 2416
      :::: :::::
1273 aIleMetAlaGlyAlaIleLeuThrGly.....ThrAlaValAlaAsnA 1288
      :::: :::::
2417 CCCAAGCAACATTAAATCAAGCCACA..... 2442
      :::: :::::
1288 sNgluGlyThrTyrPAsnLeuGlySerSerSerGlyGlnLysAsnAsnThrGly 1304
      :::: :::::
2443 .....TTAAGCGCAACACATCGCTTCGAGCAATGCTTCATTAA 2483
      :::: :::::
```

```
1305 MetLeuGluValAsnAsnSerAlaPheAsnAsnArgGlyGluPheI 1321
      :::: :::::
2484 TCTAACCAAC.....AACCGGTACAA...AACGCGACTGTGACGCTTT 2524
      :::: :::::
1321 eLeuAspAsnAspLysAsnAlaValHisIleAsnGlnSerGlyThrLeuT 1338
      :::: :::::
2525 CGGACAGCGCTAAGCAACAGTAAGCATTCGCACTCAACGCAATGTC 2574
      :::: :::::
1338 yAsnThrGlyHisMetAsnIleSerAsnSerSerHisAsnGlyAlaVal 1354
      :::: :::::
2575 TCCCTA..... 2580
      :::: :::::
1355 AsnMetTyrGlyLysAsnGlyArgPheIleAsnAspGlyThrIleAspVa 1371
      :::: :::::
2581 .....GCCGATTAAGGCAG 2593
      :::: :::::
1371 lSerAlaLysSerLeuValValSerAlaAsnAsnAlaGlyAspGlnAsnA 1388
      :::: :::::
2594 TATTCATTTTGAAGAACCGCGCTTACCGGAAAAATCAGCGCGGCAAG 2643
      :::: :::::
1388 lAspPheTyrPAsnGlnAsp.....AsnGlyValIleAsnPheAsnHis 1402
      :::: :::::
2644 GATACGCAATTACCTTAA..... 2664
      :::: :::::
1403 AspSerAlaSerAlaValLysValThrHisSerAsnPheIleAlaGlnAs 1419
      :::: :::::
2665 .GACAGCAATGAGCGCGCGCGCGGACGCAATTA.....G 2701
      :::: :::::
1419 nAspGlyIleMetAsnIle...SerGlyThrGlyAlaValAlaMetGlnG 1435
      :::: :::::
2702 GCATTTTAAC.....CTGACACAGCGCCGCTTACGCAATTC 2742
      :::: :::::
1435 lYAspLysAsnAlaGlnLeuValLysAsnGlyThrIleAsnLeuGlyThr 1451
      :::: :::::
2743 GCCTATGACACGATGCGCA.....GGCGCCAAACCGGACGCGCG 2786
      :::: :::::
1452 AlaGlyThrThrAspThrGlyMetIleGlyMetGlnLeuAspAlaAsnAl 1468
      :::: :::::
2787 AGATCGCGCGCGCGCGCGCTTCCGCGCTTCATTATTCGTTACCGCCG 2836
      :::: :::::
1468 aThrAlaAspAlaValIleGluAsnAsnGlyThrIleAsnIlePheAlaA 1485
      :::: :::::
2837 CACTTCGGCAGATCCGTTTCAACAGCGCTGACGTTAAGCGCAATTG 2886
      :::: :::::
1485 sNAspSer.....PheAlaPheSerValLeuGlyThrValGlyHisVal 1499
      :::: :::::
2887 AACGTCAGGGAACATTCGCTTATATGCGAACTCTTCGCTACCGCAG 2936
      :::: :::::
1500 ValAsnAsnGlyThrValIleAlaAspGlyValThrGly.....Se 1514
      :::: :::::
2937 CGGCAAAATTGACGTGCGGAA..... 2958
      :::: :::::
1514 rGlyLeuIleLysGlnGlyAspSerIleAsnValGluGlyMetAsnGlyA 1531
      :::: :::::
2959 .....AGTTCGGA.....GGCACTTACACCTTGCGGTGC 2988
      :::: :::::
1531 sNAsnGlyAsnSerSerGluValHisTyrGlyAspTyrThrLeuPro... 1546
      :::: :::::
2989 AACCAATACCGGCAACGACCGTAAGTCTGCGCAATTTAGCGTAGTGA 3038
      :::: :::::
1547 .....AspValProLysProAsnThrValSerValThrSe 1558
      :::: :::::
3039 AGGAAAGACACACACCGCTGTCC...GAAATCTTAATTCAACCTGC 3085
      :::: :::::
1558 rGlySerAspGluAlaGlySerMetAsnAsnLeuAsnGlyTyrValVal 1575
      :::: :::::
3086 AAAAGGAACAGCTGATCGCGCGCATGCGTTATCAGTTATCCGC... 3132
      :::: :::::
1575 aGlyThrAsnValAsnGlySerAlaGlyLysLeuLysValAsnAsnAla 1591
      :::: :::::
3133 .....AAGA 3137
      :::: :::::
1592 SerMetAsnGlyValGluIleAsnThrGlyPheThrAlaGlyThrAlaAs 1608
      :::: :::::
```







221 ArgGlnIleValArgAlaGluGlyThrAlaAsnThrValValTyr.. 236  
1584 CGGGTTGGGGGGGAGGAGTTGGATTGAAAGGGGCAATGGCTTGGTCC 1633  
237 .....AlaGlyGlyAspGlnThrValHisGlyHisAlaLeu..... 248  
1634 ACCGCATTTAAATACCAGATGAGGGCGATGATGTCMAAGCCAAATCA 1683  
249 .....AspThrThrLeuAsnGlyGlyTyrGlnTyrValHisAspGly 262  
1684 GAAAGAAGATCCACCGCTTACCATTTACAGGCAATAAAGATATTACTCAAC 1733  
263 GlyThrAlaSerAspThrValVal.....AsnSerAsp..... 273  
1734 CGGCATTAACACAACTTGATGCAAAAAAGAAATGGCTTACACAGGTT 1783  
274 .....GlyTyrGlnIleValLysAsnGlyG 282  
1784 GATTGGCGAAGAAAGATGCAACCAACAAAGCGGGGCTCATCTGAAT 1833  
282 LysValAlaGlyAsnThrThrValAsnGlnLysGlyArgLeuGlnValAsp 298  
1834 TACCAACCGGAAGAACCGGATCGCACTTACTCTTCCGGCGAACA 1883  
299 .....AlaGlyGlyThrAl 303  
1884 TTAAACGGCAATATCAGCAAAACAAAGCAATGTTTTCAGCGGGA 1933  
303 aThrAsnValThrLeuLysGln...GlyGlyAlaLeuValThrSerTha 319  
1934 GACCGACCGGACCGCTCAATCATTTAGAAAGCGGTGTCAAAAATG 1983  
319 AlaIaThrValThrGlyIleAsnArgLeuGlyAla...PheSerValVal 334  
1984 GAAGATATCCCAAGAGGAATCGTGTGGGACACAGATGGATGCAC.. 2031  
335 GlnGlyLysAlaAspAsnValValLeuGlnAsnGlyGlyArgLeuAspVa 351  
2032 ....CGCACATTTAAAGCGAAACTTCATATTCAGGGCGGACAAAGCG 2077  
351 IleuThrGlyHisThrAlaThrAsnThrArgValAspAspGlyGlyThrL 368  
2078 TGGTTTCCCGCAAT.....GTGGCAAGTGGAAGGC 2109  
368 euAspValArgAsnGlyGlyThrAlaThrThrValSerMetGlyAsnGly 384  
2110 GATTGCATTTAAGCAATCAGCCCAAGATTTCCGTTGCGACCGCA 2159  
385 GlyValLeuLeuAlaAspSerGlyAlaAlaValSerGly..... 397  
2160 TCAAAAGCCACAAATCTGTACAGTTTCGAC..... 2190  
398 .....ThrArgSerAspGlyLysAlaPheSerIleG 408  
2191 .....TGAGCGGCTGACA 2205  
408 LysGlyGlyGlnAlaAspAlaLeuMetLeuGlnLysGlySerSerPheThr 424  
2206 AGTTGTACGAAAGAAACCATTCAGACGATTAAGATTCCTTCATTCAG 2255  
425 LeuAsnAlaGlyAspThrAlaThrAspThrThrValAsnGlyGlyLeuPh 441  
2256 CAAGAC.....GACATCAGAGCAATGTCAGCTTGCCGATCAG 2296  
441 eThrAlaArgGlyGlyThrLeuAlaGlyThrThrThrLeuAsnAsnGlyA 458  
2297 CTGATTAAATCTCACAGACTGCG..... 2322  
458 IaIleuThrLeuSerGlyLysThrValAsnAspThrLeuThrIle 474  
2323 .....ACACTCAACGCGCA 2336  
475 ArgGlnGlyAspAlaLeuLeuGlnGlyGlySerLeuThrGlyAsnGlySe 491

2337 TCTTAGTGAGCGGAGACACGCACTATACGTTTACGGCAAGCCACCC 2386  
491 rValGlnLysSerGlySerGlyThrLeuThrValSer...AsnThrThrL 507  
2387 AAAAGGCAACCTCAGCTCGTGGGCAATGCCAAGCAACTTAAATCA 2436  
507 euThrGlnLysAlaValAsnLeuAsnGlnGlyThrLeuThrLeuAsnAsp 523  
2437 GCCCATTTAAACGGCAACATCGGCTCGGCAATGCTCATTTAAATCT 2486  
524 SerThrValThrThrAspValIleAlaGlnArgGlyThrAlaLeuLysLe 540  
2487 AAGCAACACCGCGTCAAAACGGCATCTG.....ACGC 2521  
540 uThrGlySerThrValLeuAsnGlyAlaIleAspProThrAsnValThrL 557  
2522 TTTCGACACGCTTAAGCAAAACGTAAGCCATTCCGCACTAACGGCAAT 2571  
557 euAlaSerGlyAlaThrTrpAsnIleProAspAsnAlaThrValGlnSer 573  
2572 GTC.....TCCCTAGCCGATAGGCAAGTATTCATTTGAAACAGCG 2615  
574 ValValAspAspLeuSerHisAlaGlyGlnIleHisPheThrSerThrAt 590  
2616 CTTTACCGGAAATCAGCGCGGCAAGATACGCAATTACATTAAAG 2665  
590 g...ThrGlyLysPhe..... 594  
2666 ACAGGAATGAGACGCTCGCTGGGCAAGCAATTAAGCAATTAACCTT 2715  
595 .....ValProAlaThrLeuLysValLysAsnLeuAsnGly 606  
2716 GACAACGCCACCATTTACATCATTCGCGCTATGACACGATGCGGACG 2765  
607 GlnAsnGlyThrIleSerLeuArg..... 614  
2766 CGCGAAACCGGCAAGTGGGCAAGATGCGCGCGCGCTTGGCGCGCTT 2815  
614 ..... 614  
2816 CCTATTATCCGTTACGCCGCCCACTTCGGCAGAAATCCGTTCAACAGC 2865  
615 .....ValArgProAspMetAlaGlnAsnAsnAlaAspArg 626  
2866 CTGACGGTAAAGCGCAATTGACGTCAGGAAACATTCGCTTATATGC 2915  
627 LeuValIleAspGly..... 631  
2916 GGAACCTCTGGCTACCGGACGGGCAAA.....TTGAACCTGGCGAAA 2959  
632 .....GlyArgAlaThrGlyLysThrIleLeuAsnLeuValAsnA 645  
2960 GTTCGGAAGCACTTACACCTTGCGTGTCAACATTAACGCAAGAACCC 3009  
645 lAsGlyAsnSerAlaSerGlyLeuAlaThrSerGlyLysGly..... 658  
3010 GTAAGCTCGAGCAATTGACGCTAGTGAAGAAAGCAACACACCGCT 3059  
659 .....IleGlnValValGlnAlaIleAsnGlyAlaThrThr 670  
3060 GTCCGAAATCTTAATTCACCCCTGCAAAAGCAACAGCTGATCGCGCG 3109  
670 rGlnGlnGlyAla.....PheValGlnGlyAsnArgLeuGlnAlaGlyA 685  
3110 CAGGCGTTATCAGCTT.....ATCCGC 3132  
685 lApeAsnTyrSerLeuAsnArgAspSerAspLusertTpyrLeuArg 701  
3133 AAAGAGCGGAGTTCCGCTGCAATATCCGTTCAAGACAAAGAGCTTTC 3182  
702 SerGlnAsnAlaTyrArgAlaGlnValProLeuTyrAlaSerMetLeuTh 718

```

3183 CGACAACTCGGCAAGCGGAGAAACAGAGCGCGCTTGACGGCAAAAC 3232
718 T..... 718
3233 AGGCAACTTGGCCGCAACAGCGGCAAAAAAGACACACGCGCAAGC 3282
719 .....GlnAlaMetasp 722
3283 CTGACGGCGGTGATTGGCGCGCGCAATGCCACCGAAAGGCAAG 3332
723 TyrAspArgIleValAlaIleSerArg..... 731
3333 TGTTCGCAACCGCGCGCGGAGCGAGC.....GGGAAATTCGCGCA 3376
732 .....SerHisGlnThrGlyValAsnGlyIleHisAsnSerV 744
3377 TTATCGAGCGGAGAGAGAAAAACGGTGCGAGCGGATTAAGACACC 3426
744 Al...ArgLeuSerIleGlnGlyIleHisLeuGlyHisAspAsnGly 759
3427 GCCTTGGGAAACAGCGGAGCGGAGCGCGCGGCTACCGCGCTT 3476
760 GlyIleAlaArgGlyAlaThrProGlnSer...SerGlySerTyrGlyPh 775
3477 CCCCCCGCGCGCGCGCGCGGATTCGCGCAACCGCACCGCCCAAC 3526
775 eValArgLeu..... 778
3527 CGCAACCCCAACCGCGCGCGCGCTGATCAGCGCTTATGCCAATAGCGCT 3576
779 .....GlnGlyAspLeuMet...ArgThrGluValAlaIleGly 789
3577 TTGAGT.....GAATTTCCGCAAGCGTCAACAGCGTTTTCG 3614
790 MetSerValThrAlaGlyValTyrGlyAlaAlaGlyHisSerSerValAs 806
3615 CGTACAGCAAGATTTGAACCGCGCTTTCGCAAGACCGCGCGCAACGCG 3664
806 PAllyAspAspAspArgIleSerArgAlaGlyThrValArgAspAspAlaG 823
3665 TTTCGACAGCGCGCATCCGCGGACACCAAACTACCGTTG..... 3705
823 LysLeuGlyGlyTyrLeuAsnLeuValHisThrSerSerLeuTyrLeuThr 839
3706 .....CAAGATTTCCGCGCGCTTACCGCA 3728
840 AlaAspIleValAlaGlnGlyThrArgHisSerMetLysAlaSerSerAs 856
3729 ACAAAACGACCTGGCGCAAAATCGTATGCAGAAAACTCGGACGCGGCG 3778
856 PAsnAspAspPheArgAlaArgIleTyrPoleTyr...LeuGlySerLeuG 872
3779 GGGTGGGACCTGTTTTCG..... 3798
872 LInHleIleuPhePheSerIleThrAspAsnLeuMetLeuGlnPheGln 888
3799 .....CACAAACGCGGAAACACTTGCAGACGCGGATCGGCACTC 3842
889 LeuGlnTyrThrTrpGlnGlyLeuSerLeuAspArgIlyAspAsnAl 905
3843 GGCACGCGCTTGGCGCGCGCTTTCGCAATACGCGCAACGCGGACGCT 3892
905 ArgIlyTyrValLysPheGlyHisGlySerAlaGlnHisValAlaArgAlaGlyP 922
3893 TCGACATTCGCG.....ATCAGCGCGGCGCGGAGTTTGTAGTAC 3930
922 HeArgLeuGlySerHisAsnAspMetThrPheGlyGlnGlyHisSerG 938
939 ArgAlaProLeuAlaArgSerAlaLysHisSerValSerGlyLeuProVa 955
3978 GCATTTAGCGCATTCAGGCAAGATACCGCGCGGATTTTCGCGGATTCGCA 4027

```

```

955 LAsnTrpTyrValGlnPro..... 961
4028 TCGAACCGGACATTCGGCGCAACGCGCTATTGTCGCAAAAGCGATTAC 4077
962 .....SerValIleArgThrPheSerSerArgIlyAspMet 973
4078 CGATACGAAACGTCATATTCGCGCACCGCGCGCTTGCATTCACCGCTA 4127
974 Arg.....ValGlyThrSerThrAlaGly..... 981
4128 CCGCGCGGCGATTAAAGCAGATTATTCATTCAAACCGCGCGCAACATTT 4177
982 .....SerGlyMetThrPheSerProSerGlnAspGlyT 993
4178 CCATACGCGCTTATTGAGCGCTGTCTATACGATCCGCTTCGCGCAAA 4227
993 hr.....SerLeuAspLeuGlnAlaGlyLeuGlnAlaArg 1004
4228 GTCCGAAACGCGCGTCATATACCGCGCGCTATTCGCGCAGATTTCCGCAAAAC 4277
1005 ValArgGlnAsnIleThrIleGlyValAlaIle..... 1015
4278 CCGCAGTTCGCGAATGGGCGCTAAACCGCGCAATCAAGTTTCACGCTGT 4327
1016 .....GlyTyrAlaHis 1020
4328 CCGTCACGCTGCGCGCGCGCAAGGG 4353
1020 eValIleSerGlySerAlaGlnGly 1028

seq_name: SwissProt_40: OMPB_RICJA
seq_documentation_block:
ID OMPB_RICJA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (OMP-B)
DE (OMP B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TH;
RC Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica."
RT Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILIARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILIARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILIARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC

```

DR EMBL: AB003681; BAA20138.1; -  
 DR InterPro: IPR003858; COMPACT-OMP.  
 DR Pfam: PF02708; COMPACT-OMP; 1.  
 KW Antigen: S-layer; Cell wall.  
 FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.  
 FT DOMAIN 528 533 POLY-GLY.  
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD599F CRC64;

## alignment\_scores:

Quality: 237.50 Length: 1561  
 Ratio: 0.343 Gaps: 77  
 Percent Similarity: 44.331 Percent Identity: 19.859

## alignment\_block:

US-09-303-518D-653 x OMPB\_RICJA ..

Align seg 1/1 to: OMPB\_RICJA from: 1 to: 1656

```

277 TTTTCTGTGTAATCGCGTAACGGCGTGGCGCA..... 309
    ||| ||| ||||| ||||| |||||
334 PhevalValSerValaspasnGlyAlaAlaThrIleaspGlyGlnVa 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
310 .TTGGCGGCGCATCAATATATGTGACGGTGCACATAACGGCGCTATA 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 IlyrAlaLysAspMetValIleGlnSerAlaAsnIleaspGlyGlnVala 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 AC.....AATGTGATTTGGTGGCGGAGGAAGCAATCCGATCAG 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 snPheArgHisIleValaspValGlyIleaspGlyThrThrAlaPheLys 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 CACCGCTTTCTTACCAATGTGAAAGAAATATTAAGCAGCAGC 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 ThrAlaIleSerIleValAlaIleThrGlnAsnSerAsnPheGlyThrTh 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 TTAACGCCATCTTATGGCGGCGATTCATATGCCGCTTTCACCAAT 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 rAsp.....PheGly.....AsnLeuAlaIleAla 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 TTGTCACAGATGACGACCTGTTGAGATGACCATATATGATGATGGG 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 IValThrValProAspThrMetThrLeuThrGlyAsnPheThrGly... 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 AAATACGCTGATTAAATAAATACCCGTATCGTTCGAATCGAGCAGG 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
423 ..... 423
600 CAGACAAATRTTGGCGTCTGATGAAGACGAAACCAATACCGCAAGTT 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 .....AspAlaAsnAsnProGlyAsn..... 430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
650 CATATCATATTTGCAAGCGCATATCTTGGCTGGGCGCAATACCTTT 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 .....ThrAlaGlyValIleThrPhe 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
700 GCACAAAATGA.....TCAGTGGTGGCAGACGTAATAGTAGTAG 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 AlaAlaAsnGlyThrLeuAlaSerAlaSerAlaAsnValAlaVa 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
741 CGAAAAA.....ATTAAACATAGCCCATATGCTTTTACCAA 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 IThrAsnAsnIleThrAlaIleGlyAlaSerGlyValGlyValGln 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
779 CAGGAGGCTCA.....TTGGGACAGCGTGGCGCACCA 810
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
471 euSerGlyThrHisThrAlaGluLeuAlaGlyAsnAlaGlySerVal 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
811 ATGTTATCTATGATGCCAA.....AAGCAAAATG 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 PheLysLeuAlaAspGlyThrValIleAsnGlyValAsnGlnThrVa 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
843 GTTATTAATGGGCTATTCGAAACAGCG..... 870

```

```

||||| ||||| ||||| ||||| |||||
504 ILeuValIleGlyValIleAlaAlaGlyAlaIleThrLeuAspGlySerA 521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
871 .....AACCCCTATATAGAAAAGCAATGGCTTCACGATGTCGT 912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 IaThrIleThrGlyAspIleGlyAsnGlyGlyGlyAlaAlaLeuGln 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
913 AAAGATTGCTCTATGATGAA.....ATCTTGCTGGAGATAC 950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
538 SerIleThrLeuAlaAsnAspAlaThrLysThrLeuThrLeuGlyGlyAl 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
951 CCATTCAGTATTTCTACGACACATCAATCAAAATGGAATACCTTTTACG 1000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
554 AsnIleIle.....SerAlaAsnGlyGlyThrIleAsnPheG 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 ACAATATATATGCGCGCAGAAAAATCGATGCCAAACATTAACACTATCT 1050
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 IAlaIleAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnIleVal 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1051 CTACCTTATGATTTAAAAACAGAACCGTTCATTTGTTAATGTTCTTT 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 ValAspCysAspLeuAlaIleAlaThrAspIleThrGlyValAlaSpAl 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1101 ATCCGAGACAGCAGACGACCTGTTATCATGCTGCAGTGGGCTC.... 1146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 AserSerLeuThrAsnAlaGlnThrLeuThrIleSerGlyThrIleGly 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 .....ACAGTTATCGACCCAGACTGAATATGAGAAAT 1182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
617 IeIleGlyAlaAsnAsnThrThrLeuGlyGlnPheAsnIleGlySerSer 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1183 ATTCCCTTATTTGACAAAGGAAA.....GCTGAATGATACTTAC 1223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
634 LysThrThrLeuAsnGlyGlyAsnValAlaIleAsnIleuValIleG 650
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1224 CAGCAACATCAAC...CAAGCGCGCGGCTTGTATTT 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
650 YAsnAsnGlySerValGlnPheAlaHisAsnThrTyLeuIleThrArg 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1261 .....GAGGCTAATTTACGGCTCGCCT...MAAAC 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
667 hrThrAsnAlaAlaGlyGlnGlyLysIleIlePheAsnProValAlaSn 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1291 AACGAACGTGGCGCAGGCGGCGTTCATATGATGATGCGATACCGT 1340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 AsnAsnThrThrLeuAlaAlaGlyThrAsnLeuGlySerAlaAlaAsnPr 700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1341 TACTTGGAAAATTAAC.....GGCGTGGCAACGACCGCTGT 1378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
700 OleuAlaGluIleAsnPheGlySerLysGlyAlaArgAlaAspThrVal 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1379 CCAAAATCGCGCAAGCAGCGCTGCTG.....GTTCAAGCCAAA 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
717 euAsnValIleGlyGlnGlyValAsnLeuTyAlaThrAsnIleThrThr 733
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1417 GGGGAAACCAAGGCTCG.....GTCAGCGTGGGCGAGCGTAAAGCAT 1460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
734 AspAlaAsnValGlySerPheValPheAsnAlaGlyGlyLysAsnIleVa 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1461 CTTAGATCAGCAGGCGGACGATCAAGCAAAAAACAGCCCTTATGTA 1510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
750 IserGlyThrValGlyGlyGlnGlnGlyAsnLys.....PheAsnThrV 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1511 TTGGCTGTGTCAGCGCAGGGGAGCGTGCATGATGCGGTAATCAG 1560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
765 AlaIleuAspAsnGlyThr...ThrValLysPheLeuGlyAsnAlaThr 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1561 TTCACCCCGCAAAACGCTATTTGGCTTTCGCGGCGAGCGTGGATTT 1610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 PheAsnGlyAsnThrThrIleAlaIle.....AsnSerThrLeuGlnI 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1611 GAACGGG...CATTCGCTTTCGTCACCGC..... 1638
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

795 eserGlyAsnTyrThrAlaAspPheIleAlaSerAlaAspGlyThrGlyI 812  
1639 ..ATTCAAAATACCGATGAGGGCGATGATTGTCACACCAATCAAGAC 1686  
812 leValIGluPheValAsnThrGlyProIleAsnValThrLeuAsnGlyGln 828  
1687 .....AAGAAATCCACGCTTACCATTTACAGGCAA 1715  
829 AlaValProValAsnAlaLeuGlyGlnIleThrValSerGlyProGlyAs 845  
1716 TAAAGATATTAATAACCGCAATTAACACACTTGATAGCAAAAAAG 1765  
845 nValValAlaAsnGlnIleGlyAsnAlaGlyAsn..... 856  
1766 AAATTCCTACACGGTTGGTGGCGAGAAAGATGACACCAAAAGAC 1815  
857 .....TyrHisGlyAlaMetThrAsp..... 863  
1816 GGGGGGCTCATCTGAATTACCAACCGAAGACGGATCGACTTACT 1865  
864 .....ThrIleAlaPheGlnAsnSerIleuGlyAlaValLeuPhe 877  
1866 GCTTCCGCGGAAACAAATTTAAC.....GGCAAT..... 1896  
877 eleuProSerGlyIleProPheAsnAspAlaGlyAsnThrIleProLeu 894  
1897 ..ATCACCCAAACAAACGGCAACTGTTTTCAGCGGACGACACCG 1944  
894 hTleLysSerThrValGlyAsnGlnuThrAlaGluGlyPheSerValPro 910  
1945 CACGCTTCATATTCATTTAGAACGGGGTGTCAAAAATGAAAGTATCCC 1994  
911 SerVal.....IleValSerGlyValAspSerVal.....IleAl 922  
1995 ACAAGAGAAATCGTGTGGACACGATGATC..... 2028  
922 aAspGlyGlnValIleGlyAspGlnAsnAsnIleValGlyLeuGlyLeuG 939  
2029 .....GACCGACATTTAAACCGGAAACCTCCATATTCAGGCGGACAA 2073  
939 lySerAspAsnGlyIleIleValAsnAlaThrThrLeuTyrAlaGlyLe 955  
2074 GCGGTGTTTCCCGCATGTTGCCAAAGTGAAGGCGATGCGATTAG 2123  
956 GlyThrIleAsnAsnGlnGlyThrValThrLeuSerGlyGlyValPr 972  
2124 CAATCAGCCCAAGCAGTTTCGGTGC.....GCAC 2155  
972 oAsnThrProGlyThrValTyrGlyLeuGlyThrGlyIleGlyAlaSerL 989  
2156 CGCATCAAAAGCCACAAATCTGTACACGCTCGACTGACGGGTGACA 2205  
989 yspheLysGlnValThrPheThrThrAspTyrAsnLeuGlyAsnIle 1005  
2206 AGTTTACCGAAAAACATTAACGACATTAAGATGATGCTCA..... 2250  
1006 IleAlaThrAsnThrThrIleAsnAspGlyValThrValThrThrGlyG 1022  
2251 .....TTGAGCAAGACGACATTCAGGACGACGATGACAC 2284  
1022 yIleAlaAlaGlyGlyIleAlaGlyThrAspPheAspGlyLysIleThr 1038  
2285 TTGCGGATCAGCGCTTAATCTCAGAGACTTGCCACATCAACGCG 2334  
1039 .....LeuGlySerValAsnGly 1044  
2335 AATCTTAGTGA.....GGCGAGACGACGACTA 2363  
1045 AsnAlaAsnValArgPheAlaAspGlyIlePheSerAsnSerThrSerMe 1061  
2364 TACGGTTACGGCAGCAGCCAAACGCAACTCAGCCGCGGGACA 2413  
1061 tIleValThrThrLysAla..AsnAsnGlyThrValThrTyrIleuGlyA 1077  
  
2414 ATGCC..... 2418  
1077 snAlaPheValGlyAsnIleGlyAspSerAspThrProValAlaSerVal 1093  
2419 .....CAAGCAACATTTAATCAGCCACATTTAAACGGCAAC.....AC 2456  
1094 ArgPheThrGlySerAsnAsnGlyAlaGlyLeuLysGlyAsnIleTyrSe 1110  
2457 ATCGGCTTGCGACATGCTTCATTTAATCTA.....AGCAACACG 2497  
1110 rGlnValIleAspPheGlyThrTyrAsnLeuGlyIleValAsnSerAsnV 1127  
2498 CCGTACAAACGCGACAGTCTGACGCTTCCGACACGCTAAGCAAAAGTA 2547  
1127 alIleLeuGlySerThrThr..... 1134  
2548 AGCCATTCGCACTACAGCGCAATGCTCTCAGCCGATAGCGATATT 2597  
1135 .....AlaIleAsnGlyLysIleAspLeuThrAsn..... 1145  
2598 CCATTTTGAACAGCGCGCTTACCGGAAATACAGCGGCGCAGAGATA 2647  
1146 .....ThrIleuThrPheAlaGlyGlyThrSerT 1155  
2648 CGGCTTACACTTAAAGACAGAGATGACGCGCTGCGGACAGGAA 2697  
1155 hTrrpGlyAsnAsnThrSerIleGluThrThrLeu.....Thr 1167  
2698 TTAGGCAATTTAACTTCGACACAGCGACCATTTACCTCAATTCGGCCTA 2747  
1168 LeuAlaAsnGlyAsnIleGlyHisIleValIle..... 1178  
2748 TCGACAGATGCGCGAGCGGCAACCGCACTGCGGAGATGCGCCGC 2797  
1179 .....AlaGluGlyAlaGlnValAsnAla..... 1186  
2798 GCCGCGCTTCCGCGCGCTCCCTATTAATCCGTTACGCCGCACTTCGGCA 2847  
1187 .....ThrThrThrGlyThrThrThrIleAsnValGlnAspAsnAla 1200  
2848 GAATCCGCTTCAAC.....ACGCTGACGCTA..... 2874  
1201 AsnAlaAsnPheSerGlyThrGlnThrTyrThrLeuIleGlnGlyVal 1217  
2875 .....AACGGCAATTTGAACGT.....CAGGACAT 2902  
1217 aArgPheAsnGlyThrLeuGlyGlyProAsnPheThrValThrGlySerA 1234  
2903 TCCGCTTTATGTCGGAACCTTCCGCTACCGCAGCGCAATTTGAAGCTG 2952  
1234 snArgPheValAsn.....TyrGlyLeuIleAlaGala 1244  
2953 GCG.....GAAAGTTCGAAGGCACTTA 2975  
1245 AlaAsnGlnAspTyrValIleThrArgThrAsnAlaGlnAsnIleVal 1261  
2976 CACCTTGGCTGTCAACATAC.....GGCAAGAACCCGTAAGTCTCG 3019  
1261 lThrAsnAspIleThrAsnSerProPheGlyAlaProGlyValGlyG 1278  
3020 AGCAATTGACGGTAGTGAAGAAAGACACACACCGCTGTCCGAAAT 3069  
1278 lAsnValThrThrPheValAlaAlaThrAsnThrAlaLaryAsnAsn 1294  
3070 CTTAATTTCAACCCGCAAAACGAACAGCTGATGCGGCGCATGGCTTA 3119  
1295 Leu..... 1295  
3120 TCAGCTTATCCGCAAGACGCGAGTTCGCTGATATTCGGGTCAAG 3169  
1296 .....LeuLeuAlaLysAsnSer..... 1301



```

RN [4]
RP PARTIAL SEQUENCE.
RC STRAIN=BREINL;
RA Ching W.M., Carl M., Dasch G.A.;
RT Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.
RL Mol. Immunol. 29:95-105(1992).
RN [5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
RT Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing.
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37647; AAA26390.1; ALT_INIT.
DR EMBL: AF161079; AAD42234.1; -.
DR EMBL: AJ235273; CA15140.1; -.
DR InterPro: IPR003858; OMPA_OMP.
DR Pfam: PF02708; OMPA_OMP.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 V->A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y->D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A->S (IN STRAIN BREINL).
FT CONFLICT 178 179 AA->VC (IN REF. 1).
FT CONFLICT 191 201 TQEAFLTLGA->INSRSSYHLVS (IN REF. 1).
FT CONFLICT 212 212 T->I (IN REF. 1).
FT CONFLICT 313 313 Q->L (IN REF. 1).
FT CONFLICT 1104 1104 D->G (IN REF. 2).
FT CONFLICT 1123 1123 T->S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Alignment scores:
Quality: 237.00 Length: 1461
Ratio: 0.362 Gaps: 68
Percent Similarity: 44.832 Percent Identity: 19.233

Alignment_block:
US-09-303-518d-653 x OMPB_RICPR ..
Align seg 1/1 to: OMPB_RICPR from: 1 to: 1643

661 GGAAGCGCATTTCTGGCTGCTGGTCATACCTTGCCACAAATGG 710
|||||:.....:|||||:|||||:|||||:|||||:
250 AIAAsnTThrLeuAsnLeuGlnValGlyGlyAsnThrIleAsnPhAsnG1 266
711 A...TCAGGTGGTGCACAGTCACTAGTAGTGCAGAAAAATTAACATA 757
|:|||||:|||||:|||||:|||||:|||||:
266 ylleAsnpglyThrGlyLysLeuValSerLysAsnGlyAlaIat 283
758 GCGCATATGGTGTGTTTACCACACA...GGAGCTCATTTGGCGCAGAGTGC 804
:::.....:|||||:|||||:|||||:

```

```

283 hrcGluPhAsnValThrGlyThrLeuGlyGlyAsn..... 294
805 TCACCAATGTTATCTATGATGCCCAAGCAAGGTGTTAATAATGG 854
295 .....:|||||:|||||:|||||:|||||:
855 GGTATTCGAA.....LeuYsG1 297
297 ylleIleGluLeuAsnThrAlaValAlaGlyLysLeuIleSerGln 314
865 .....ACAGGCAACCCCTATATGCAAAAAGCAATGCTCCAGCTAGT 909
314 IyGlyAlaIleAsnAlaValIleGlyThrAspAsnGlyAlaGlyrrGla 330
910 CGTAAGATTGTTCTATGATGAATCTTGCTGGAGATACCATTCAGT 959
331 .....AlaGlyPheIleValSerVa 337
960 ATTCTACGAACACATCAAAATGGAAATGCTTTTAAAGCAATATA 1009
337 .....:|||||:|||||:|||||:|||||:
1010 ATGGCGCA.....GAAAAATGATGCCCAACATAACACTATTC 1050
341 snAlaIatThrIleSerGlyGlnValTyrAlaLysAsn..... 353
1051 CTACCTTATGATTAAACACAGACCGTTCATGTTAATGTTCTTT 1100
354 .....MetValIleGlnSerAlaAsnAlaGlyL 363
1101 ATCCGAGACAGACAGAAACCTGTTATATGCTGAGTGGGGTCACA 1150
363 yGlnValThrPheGluHisIleValAspValGlyLeuGlyThrThra 380
1151 GTTATGACACCCACAGCTGAATTAATGAGAAATTTCTTATATGACANA 1200
380 snPheLys.....Thr 383
1201 GGAAGAGTGAATGTACTTATACCAACATCAACCAAGGCGCG..... 1245
384 AlaSperLysValIleIleThrGluAsnSerAsnPheGlySerThrAs 400
1246 .....GGCGGTTTG.....TATTTG 1261
400 nPheGlyAsnLeuAspThrGlnIleValValProAspThrLysIleLeu 417
1262 AGGCTATTTTACGCTGCTGCTAAACACGAA..... 1296
417 ysglyAsnPhnIleGlyAspValLysAsnGlyAsnThrAlaGlyVal 433
1297 ...ACGTGGCAGAGCGCGCGCTTCATATCACT.....GATGGCAG 1334
434 IleThrPheAsnAlaAsnGlyAlaLeuValSerAlaSerThrAspProAs 450
1335 TACCGTTACTTGGAACTAAACGCG..... 1359
450 nIleAlaValThrAsnIleAsnAlaIleGluAlaGlyValAlaGlyVal 467
1360 .....GTGGCAACGACCGCTGTC..... 1380
467 alGluLeuSerGlyIleHisIleIleAlaLeuAsnThrGlyLeuGlyAsn 483
484 SerIlePheLysLeuAlaAspGlyThrValIleAsnGlyProValAsnG1 500
1422 AAC.....CAAGCTGGGTAGCGTG 1444
500 nAsnAlaLeuMetAsnAsnAlaLeuAlaIaGlySerIleGlnLeu 516
1445 GCGAGCTAAAGTCATCTTAGTCAGCAGCGGACGACGTAAGCAAAAA 1494
517 ..AspGlySerAlaIleIleThrGlyAspIleGlyAsnGlyValAsn 532

```



1495 CAAGCCTTTAGTGAATGGCGTGTGTCAGCGCGGACGGGACGGG...CA 1541  
1533 A1a1a1eUglnH1s1eThrLeuAlaAsnAlaSerLys1leUa1 549  
1542 ACTGAATGCCGATTAATCAGTTCAACCCCGAC.....AACTCATTT 1582  
549 aleuAspGlyAlaAsn1le1leGlyAlaAsnValGlyAla1leHisP 566  
1583 TCGGCTTTCGCGGCGACGTTTGGATTGTAAGCGGCGATTCGCTTCGTC 1632  
566 heGlnAlaAsnGlyThr1le..... 573  
1633 CACCGCATTCAAATATCCGATGCAAGGCGCGATGTGTCAACCAATCA 1682  
574 ..LysLeuThraSnThrGlnAsnAlaVal.....AsnPh 586  
1683 AGCAAGAAGATCCACCGTTTACCATTAAGCAATTAAGATTACTATCA 1732  
586 eAspLeuAsp1leThrThraSplyThrclyValaValaAsp1leSer1 603  
1733 CCGGCATTAACAACAAGTTGGATAGCAAAAAAAGAAATTCCTACACG 1782  
603 eutHraSnaSnGlnThrLeu.....Thr1leAsnGly 613  
1783 TGGTTTGGCGAGAAAGATGCA...ACCAAAACGAAGCGCGCGCTCATCT 1829  
614 Ser1leGlyThrValaValaAlaAsnThrLysThrLeuAlaGlnLeuAsn1l 630  
1830 GAATTATCCAAACCGGAAGAAGCGATCGCATCTTACTGCTTCGCGCGGAA 1879  
630 eGly.....SerSerLysThr1leLeuAsnAlaGlyAspV 642  
1880 CAATTTAAACGCGCATATCACGCAACAACGCGCAACTGTTTTCAGC 1929  
642 a1a1a1eAsnGlnLeuVal1leGlnAsnAsnGlySerValGlnLeuAsn 658  
1930 GCGAGA.....CCGACACCGCAGCGCTACAT..... 1956  
659 HisAsnThrTyLeu1leThrLysThr1leAsnAlaAlaAsnGlnGlyG1 675  
1956 ..... 1956  
675 n1le1leVala1a1a1aAspProLeuAsnThraSnThrThr1leuAlaSpG 692  
1957 .....CATTTAGAACGCGGCTGCTCAAAAATGAA..... 1986  
692 LyrThraSnLeuGlySer1aGlnAsnProLeuSerThr1leHisPheAla 708  
1987 .....GGTATCCCAACGAGCAAT 2006  
709 ThrLysAla1a1aAsnAlaAspSer1leLeuAsnValGlyLysGlyValaS 725  
2007 CGTGTGGGACAAGATTTGGATGACCGCAACATTTAAACGGGAAAACCTCC 2056  
725 n1euTyra1a1aAsnAla1leThrThraSnAspAlaAsnValGlySerLeuH 742  
2057 ATATTTCAG...GGCGACAAGCGGCTGTTCCCGCAATGTTGCCAAAAGT 2103  
742 1sPheArGserGlyGlyThrSer1leValaSerGlyThrValaGlyGlyGln 758  
2104 GAAGGCGATTGGCATTTAGCAATACAGCCAGCCAGGATTTTCGGGTCCG 2153  
759 GlnGly.....HisLysLeuAsnAsnLeu1leLeuAspAsnGly..... 771  
2154 ACCGCATCAAAAGCCACACAATCTGTACACGTTTGGACTGGAGGGGTGTA 2203  
772 .....ThrThraValLysPheLeuGlyAspT 780  
2204 CAAGTTGTATCCGAAAAACCATTCATTCAGCAGATTAAGTATGCTTCATG 2253  
780 hrThraPheAsnGlyGlyThrLys1leGlyGlyLysSer1leLeuGln1le 796

797 SerTrasnaSnThrThrThraSph1sValGlnSerAla1aAspAsn..... 810  
2304 AAATCTCACAGGACTTGCACACOTCAACGCGCAATCTTAGTCAGCGGAG 2353  
811 .....ThrclyThrLeuGlnPheValaSnT 819  
2354 ACACGCACTTACGTTTACGCGCAC.....GCCACCCAAAACGCGCAC 2397  
819 hrAspPro1leThrValThrLeuAsnLysGlnGlyAlaTyPheGlyVal 835  
2398 CTCAGCCCTGTG.....GGCAATGGCCAAAGCAAGCATTTAATCAAGC 2438  
836 LeuYsGlnVal1le1leSerGlyProGlnAsn1leValaPheAsnGln1l 852  
2439 C.....ACATTAACGCGCAACACATCGGCTTCG 2467  
852 eGlyAsnValaGly1leValHisGly1leAla1aAsnSer1leSerPheG 869  
2468 ACAATGCTTATTAATCAAC..... 2490  
869 lAsnAlaSerLeuGlyThraSerLeuPheLeuProSerGlyThrProLeu 885  
2491 .....AACACGCGCGCAAAACGCGACGTGACGCTTC 2525  
886 AspValLeuThr1leLysSerThraValGlyAsnGlyThraVal..... 899  
2526 CGACACGCTTAAGCAAAAGTA.....AGCCATTCGCGAC 2560  
900 .AspAsnPhaAsnAlaPro1leValaValaValaSerGly1leAspSerMet1 916  
2561 TCACGCGCATGTCTCCCTTAGCCGATTAAGCACTATTCATTTTGAAC 2610  
916 lAsnaSnGlyGln1le1leGlyAspLys..... 925  
2611 AGCGCTTACCGGAAAATCACCGCGCGCAAGATACGATACACTT 2660  
926 ..LysAsn1le1leAlaLeuSerLeuGlySerAspAsnSer1leThra 941  
2661 AAAGACAGCAATGAGAGCTGCGCGTGGGACGCAATTAAGCAATTTAA 2710  
941 lAsnAlaAsn.....ThrLeuTySerGly1leArGThrThrLysAsna 956  
2711 ACCTTGACAAAGCAGCATTAACATTCGCGCTATCGACAGAT... 2757  
956 sn.....GlnGlyThraValThrLeuSerGlyGlyePProAsnSnPro 970  
2758 .....GCGGACAGCGCGCAAAACGCGAGTGGCGGATGCGCC 2795  
971 GlyThr1leTyrcGlyLeuGlnGlyAsnGlySerPro..... 983  
2796 GCGCGCGCGTTCGCGCTTCCTATTAATTCGTTACGCGGCAACTTCG 2845  
984 .....LysLeuYsGlnValThrPheThraSPT 994  
2846 CAGAAATCCGT.....TTCACACGCTACGCGTAAC..... 2877  
994 yTraSnaSnLeuGlySer1le1leAlaAsnAsnValThr1leAsnAspTy 1010  
2877 ..... 2877  
1011 ValThrLeuThrThraGly1leAlaGlyThraSphAspAlaTyS1l 1027  
2878 .....GGCAATTTGAAGCGTACAGGAACATTCGCTTATGTGGAGAC 2920  
1027 eThrLeuGlySerValaSnGlyAsnAlaAsnValaArgPheValaSPeT 1044  
2921 TCTTCGCTACCGCAGCGCAAAATTTGAAGCTGGCGGAAGTTCCGAGGC 2970  
1044 hrPheSerAspProArgSerMet1leVala1aThrclnAlaAsnGlySgly 1066  
2971 ACTTACAC.....TTGGCTGTCAACATACCGGC.....AA 3002

```

1061 ThrValThrTyrLeuGlyAsnAlaLeuValSerAsnIleGlySerLeuAs 1077
1077 pnhProValAlaSerValArgPheThr.....GlyAsnAspSer. 1090
3053 CACCGCGTCCGAAATCTTAATTCACCCCTCAAAAGCAACACGCTGCAT 3102
1091 .....GlyAlaGlyLeuGlnGlyAsnIleTyrSerGlnAsnIleAsp 1104
3103 GCCCGCGCATGGCGTTATCAGCTTATCCGCAAGACGGCGAGTCCGCT 3152
1105 PheGlyThrTyrAsnLeuThrIle..... 1112
3153 GCATATCCGCTCAAAAGAACAGAGCTTTCCGCAAACTCGGCAAGGGCG 3202
1113 .....LeuAsnSerAsnValIleLeuGlyG 1121
3203 GAGAAACAGAGAGCGCCCTTGACGCAAAAGCAACCTGGCCGCCAA 3252
1121 LysIYThr.....ThrAlaIleAsnGlyGluIle..... 1130
3253 CAAACAGCGGGAAGACACACCGCAAGCCCTTGACGCGCTGATTGCGGC 3302
1131 .....AspLeuLeuThrAsnAsnLeuIlePheAlaAs 1141
3303 CGGCGCATGCGCACGCAAAAGCAGAAAGTGTTCGCAACCGCGCGCGC 3352
1141 nGlyThrSerThrTyrGlyAspAsnThrSerIleSerThrLeuAsnV 1158
3353 AGGACGCGGGGGAATATCGCGCATATGACCGCGAGAGAGCAAAAA 3402
1158 alSerSerGlyAsnIleGlyGlnValIleAlaGlnAspAlaGlnVal 1174
3403 CGGCGTACGCGGATAAAGACACCGCTTGCGGAAACGCGCAAGCGGA 3452
1175 AsnAlaThrThrThrGlyThrThrIleLysIleGlnAspAsnAlaAs 1191
3453 AACCGCG.....CGGCTACACGCGCTTCGCCCGCGCGCGCGCGC 3499
1191 nAlaAsnPheSerGlyThrGlnAlaThrThrLeuIleGlnGlyAla 1208
3500 GGGATTTCGCCCAACCGCAGCCCAACCGCAACCGCAGCGCGAC 3549
1208 Tg..... 1208
3550 CTGATCAGCCGTTATGCCATAGCGGTTTGAGTATTTCCGCGAGCT 3599
1209 .....PheAsnGlyThrLeuGlyAlaProAsnPheAlaValThrG 1222
3600 CAACAGCGTTTTCGCGCTACAGAGCAATGGACCGCGTGTTCGCGAG 3649
1222 YSerAsnIlePhe.....ValLysTyrGluLeu..... 1231
3650 ACCGCGCGACGCGCTTGGACAAGCGCATCCGCGACACCAAACTAC 3699
1232 .....IleArgAspSer..... 1235
3700 CGTTCGCAAGATTTCGCGGCTACCGCGCAACAAACGCTCGCCAAAT 3749
1236 .....AsnGlnAspTyr..... 1239
3750 CGGTATGCGAAGAAACCTCGCGACGCGCGCGTGCATCCTTTTTCG 3799
1240 .....ValLeuThrArgT 1244
3800 ACAACCGGACCGGAACACCTTGACAGCGCATCGCAACTCGCGACG 3849
1244 hrAsnAspValIleuAsnValValThrThrAlaValGlyAsnSerAlaIle 1260
3850 CTTGCCACAGCGTGGCTTTGGGCAATAGGCGCATGCGCA.....GGTTCGA 3896
:::|||||

```

```

1261 AlaAsnAlaProGlyVal..SerGlnAsnIleSerArgCysLeuGluSerT 1277
3897 CATCGGATCAGCGCGCGCGGCTTTTACTACG..... 3931
1277 hrAsnThrAlaAlaTyrAsnAsnMetLeuLeuAlaLysAspProSerAsp 1293
3932 ...GCAGCGTTTCAGACGCTACAGAGCAAAATTCGCCCGCGCGTGC 3978
1294 ValAlaThrPheValGlyAlaIleAlaThrAspThrSerAla..... 1308
3979 CATTCAGCATTCAGCGCAATACCGCGAGTTTCGGCGGATTCGCAT 4028
1309 ..ValThrThrValAsnLeuAsnAspThrGlnLysThrGlnAspLeuAs 1325
4029 CGAACCGCATATCGCGCAACGCGCTATTTCGTCCAAAAGCGGATTACC 4078
1325 ersnArg..LeuGlyThrLeuArgTyrLeuSerAsnAla..... 1337
4079 GATACGAAACGTCATATTCGCCACCGCGCGCTTCATTCACCGCTAC 4128
1338 .....GluThrSerAspValAla.....GlySerAlaThrGlyAlaVal 1350
4129 CGGCGGCGCATTAAGCAGATTTCATTTCAAACGCGCGCAACACATTC 4178
1351 SerSerGlyAspGlnAlaGlnValSerTyr.....GlyValThr 1363
4179 CATCAGCGCTTATTGAGCGCTGCTATACCGATGCGCGCTCGCGC.... 4224
1363 pAlaLysProPheThrAsnIleAlaGlnAspLysLysGlyAla 1380
4225 .....AAAGTCGACGCGC..... 4239
1380 laGlyThrLysAlaLysThrThrGlyValValAlaGlyLeuAspThrLeu 1396
4240 .....GTCAATACGC 4250
1397 AlaSerAspAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAs 1413
4251 CGTATTGGCGCAGATTTCGGCAAAACCGCGAGTGGGGAATGGCGGTAA 4300
1413 pLleLysHisGlnAspTyrLysLys.....GlyAspL 1424
4301 ACGCCGAAATCAAGGTTTCACGCTGCCCTCCACGCTGCCCGCGCAAG 4350
1424 ysthrAspIleAsnGlyLeuSerPheSerLeuTyrGlySerGlnGlnLeu 1440
4351 GGGCGCAATTGGAAGCGCACACGCGCG 4380
1441 ValLysAsnPhePheAlaGlnGlyAsnAla 1450
seq_name: SwissProt_40:OMPb_RICCN
seq_documentation_block:
ID OMPb_RICCN STANDARD; PRT; 1655 AA.
AC O9KKA3; O9KK98; O9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Romp B) (containing antigen) (Cell surface antigen 5) (Scab) (RompB)
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
DE OMPb OR RC1085.
GN Rickettsia conorii.
OS Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiales; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,

```

RA Roullet D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";  
 RL Science 293:2093-2098(2001).  
 RN [2].  
 RP SEQUENCE OF 33-1649 FROM N.A.  
 RC STRAIN-Indian tick typhus, and Malish 7;  
 RX MEDLINE-20393643; PubMed-10939649;  
 RA Roux V., Roullet D.;  
 RT "Phylogenetic analysis of members of the genus Rickettsia using the  
 RT gene coding the outer-membrane protein ompB (ompB)."  
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).  
 RN [3].  
 RP SEQUENCE OF 353-1655 FROM N.A.  
 RC STRAIN-Malish 7;  
 RA Stenos J., Walker D.;  
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia  
 RT australis, the most divergent Rickettsia of the spotted fever group";  
 RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC SIMILARITY).  
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIA OMPA/OMP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AE008659; AAL03623.1; -  
 DR EMBL: AF123721; AAF34124.1; -  
 DR EMBL: AF123726; AAF34129.1; -  
 DR EMBL: AF149110; AAD39533.1; -  
 DR InterPro: IPR003858; OMPA\_rOmpB.  
 DR Pfam: PF02708; rOmpA\_rOmpB; 1.  
 KW Antigen; S-layer; Cell wall; Complete proteome.  
 FT CHAIN 1 1334 120 KDA SURFACE-EXPOSED PROTEIN.  
 FT VARIANTS 1335 1655 32 KDA BETA PEPTIDE.  
 FT VARIANTS 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANTS 1139 1139 KD -> GH (IN REF. 3).  
 FT CONFLICT 353 354 E -> S (IN REF. 3).  
 FT CONFLICT 776 776 E -> D (IN REF. 3).  
 FT CONFLICT 1159 1159 G -> S (IN REF. 3).  
 FT CONFLICT 1177 1177 H -> R (IN REF. 3).  
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).  
 SO SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

alignment\_scores:  
 Quality: 237.00 Length: 1347  
 Ratio: 0.337 Gaps: 65  
 Percent Similarity: 45.443 Percent Identity: 18.293

alignment\_block:  
 US-09-303-518D-653 x OMPB\_RICCN  
 Align seg 1/1 to: OMPB\_RICCN from: 1 to: 1655

679 CTCGTCGTCGCAATACCTTTGACACAAATGATGAGTGTGTCGACAGT 728  
 :::::::::: :::::  
 :::::::::: :::::

199 IleuGlyAspAsnAlaValIleValAsnGlyAla...AsnGlyThrLe 214  
 729 CAACCTAGGTAGCGCAAAAATTAACATGACCCATGTTGGTTTTCACCA 778  
 ::::: ::::: ::::: :::::  
 214 uAsnValThrAsnGlyPheIleGlyValSerSerLysPheAlaThyV 231  
 779 CAGCAGGC...TCATTTGGCGACAGTGGCTCCACCAATGTTATCTATGAT 825  
 ::::: ::::: ::::: :::::  
 231 AlaAsnValIleAsnIleGlyAspGlyGlnGlyIleMetPheAsnThrAsp 247  
 826 GCCCAAAAGCAAAAGTGGTTA..... 846  
 ::::: ::::: ::::: :::::  
 248 AlaAspAsnValAsnThrLeuAsnLeuGlnAlaAsnGlyAlaThrIleTh 264  
 847 ATTAATGGGGTATTGCAACAGCAACCCCTATATGAGAAAGCAATG 895  
 ::::: ::::: ::::: :::::  
 264 rPheAsnGlyThrAspGlyThrGlyLeuValLeuLeuSerLysAsnA 281  
 896 GCTTCGCAAGTACGTTCCGAAAGATTGCTCTATGATGCAAACTTTGCTGGA 945  
 281 Ia.....AlaAlaThrAspPheAsnValThrGlySerLeuGlyGly 294  
 946 GATPACCATTCAGTA.....TTCAGCAACCAACATCAAAATGGCAATA 989  
 ::::: ::::: ::::: :::::  
 295 AsnLeuGlyGlyIleIleGluPheAsnThrValAlaValAsnGlyGln 311  
 990 CTTTTCATC.....GACATTAATAATGCGC 1015  
 ::::: ::::: ::::: :::::  
 311 uLysAlaAsnAlaGlyAlaAsnAlaValIleGlyThrAsnAsnGlyA 328  
 1016 CAGGAAAAATCGATGCCAACAATTAACATATTCCTACCTTATGATTA 1065  
 ::::: ::::: ::::: :::::  
 328 IaGlyTrg.....AlaAlaGlyPheVal 336  
 1066 AAAACAGCAACCGTCATGTTTAAATGTTCTTATCGAGACGCAAG 1115  
 ::::: ::::: ::::: :::::  
 337 SerValAspAsnGlyLysValAlaThrIleAspGlyGlnValTyrAla 353  
 1116 AGAACCTGTTTATCATGCTGCAGAGTGGGCTCAC.....AGTTATGCA 1159  
 ::::: ::::: ::::: :::::  
 353 SAspMetValIleGlnSerAlaAsnAlaValGlyGlnValAsnPheArg 370  
 1160 CCAGACTGATTAATGA...GAAATATTTCTTATGACAAAGGAA 1206  
 ::::: ::::: ::::: :::::  
 370 ILeIleValAspValGlyThrAspGlyThrAlaPheLysThrAla 386  
 1207 GGTGAATGATTAATCAACGACCAATCAACCAAGCGCG..... 1245  
 ::::: ::::: ::::: :::::  
 387 SerLysValAlaIleThrGlnAsnSerAsnPheGlyThrThrAspPhe 403  
 1246 .....GGCGGTTTGATTTTGAGGGTA 1267  
 403 yAsnLeuAlaIleGlnIleIleValProAsnThrMetThrLeuAsnGly 420  
 1268 ATTTACGGCTCGCCCTAAACACGAAAGCGTGGCAAGCGCGGGGT 1317  
 ::::: ::::: ::::: :::::  
 420 snPheThrGlyAspAlaSerAsnProGlyAsnThrIleAlaGly..... 433  
 1318 CATATCAGTGAATGCGCATGACGTTACTGTGAAGATAACGCGC..... 1359  
 ::::: ::::: ::::: :::::  
 434 .....ValIleThrPheAspAlaAsnGlyThrLeuAl 444  
 1360 .....GTGGCAAAAGCAGCGCTGCGCA 1381  
 444 aSerAlaSerAlaAspAlaAsnValAlaValThrAsnAsnIleThrAla 461  
 1382 AAATGGCAAAAGCAGCGCTGCTGTTCAAGCCAAAGGAGGAA..... 1422  
 ::::: ::::: ::::: :::::  
 461 IeGlnAlaSerGlyAlaGlyValValGlnLeuSerGlyThrIleAlaAla 477  
 1423 .....AACCAAGGCTCGGTC...AGCGTGGCGCAAGGTAA 1454  
 ::::: ::::: ::::: :::::  
 478 GluLeuArgLeuGlyAsnAlaGlySerValPheLysLeuAlaAspGly 494



```

1010 nAlaThrIleAsnAspGlyValThrValThrThrGlyGlyIleAlaGlyI 1027
2878 .....GCCAAATTGAACGTCGAGGAC 2901
1027 IeGlyPheAspGlyLysIleThrLeuGlySerValAsnGlyAsn 1043
2902 TTCGCGTTTATGTCGGAACCTTCGGCTACCGCAGCGCAAAATTGAAGCT 2951
1044 ValArgPheValAspGlyIleLeuSerHisSerThrSerMetIleGly 1060
2952 GCGGGAAGTTCGGAAGGACCTTACCTTGCT .....GTCA 2989
1060 rThrLysAlaAsnAsnGlyThrValThrThrLeuGlyAsnAlaPheValG 1077
2990 ACAATACCGGCAAGAA .....CCGTAAGTCGAGCAATTCAGCGGA 3033
1077 LysnIleGlyAspSerAspThrProValAlaSerValArgPheThrGly 1093
3034 GTGGAAAGAAAGAACACACACCGCTGTCGAAATCTTAATTTCACCT 3083
1094 SerAspGly .....GlyAlaGlyLeuGlnGlyAsn 1104
3084 GCAAAGCAACACACGTGATGCGCGGCATGCGCTTATCAGCTTATCCGA 3133
1104 eTyTrSerGlnValIleAspPheGlyThrTyrsnLeuGlyIleSerAsn 1120
3134 AAGACGCGCATGCTCGCTGATATCCGTCAAGACACAGCGTTTGC 3183
1121 .....Ser 1121
3184 GACAACTCGGCAAGCGGAGAAACAGAGCCGCTTGACGCAAAACA 3233
1122 AsnValIleLeuGlyGlyIleThr .....AlaIleAsnGly 1134
3234 GGCACAACTTGGCGCAACACAGCGGAAAGAACAGCGCAAGCC 3283
1135 .....LysIleAsnLeuArgThr 1141
3284 TTGACGCGTATGCGCGCGCGCAATCCACCGAAAGGACAGAAAGT 3333
1141 snThrLeuThrPheAlaSerGlyThrThrThrThrGlyAsnAsnThrSer 1157
3334 GTTGGCGAAGCGCGCGCGGAGCAGCGGGAATGCCGCGATTATGCA 3383
1158 IleGlyThrThrLeuThrLeuAlaAsnGlyAsnIleGlyAsnIleVal 1174
3384 GCGGAGGAGAGAGAAAAACGGGTCAGCGGATTAAGACACCGCTTGG 3433
1174 eLeuGlnGlyAlaGlnValAlaAsnAlaThrThrThrThrThrIle 1191
3434 CGAAACAGCGCGAGCGGAAACCGG .....CCGCTACAC 3471
1191 yValGlnAspAsnAlaAsnAlaAsnPheSerGlyThrGlnThrTyThr 1207
3472 GCGTTCCCGCGCGCGCGCGCGCGCGGATTGGCGCAACGCGACCC 3521
1208 LeuIleGlnGlyGlyAlaArgPheAsnGlyThrLeuGlyGlyPro 1222
3522 CCNACCGCAACCCCAACCGCAGCGGACCTGATC .....AGCC 3559
1223 .....AsnPheValValThrGlySerAsn 1231
3560 GTTATCCCAATAGCGGTTTGAAGTAATTTCCGCGACGCTCAGACGCTT 3609
1231 rGpPheValAsnTyGlyLeuIleArg...AlaAlaAsnGlnAspTyVal 1246
3610 TTGCGCGTACAGAGCAATTCGACCGCTGTTGCCGGAAGCCGCGCA 3659
1247 IleThrAlaGlyThrAsnAlaGlnValValThrAsnAspIleAlaAs 1263
3660 CGCGCTTGG .....ACAAGCGGCA 3679

```

```

1263 nSerSerPheGlyGlyAlaProGlyValGlyGlnAsnValThrThrPhe 1280
3680 TCCGGGACACCAACACTACCGCTTCGAAG .....ATTTCGCGCC 3720
1280 aIAsnAlaThrAsnThrAlaIleAlaTyrsnAsnLeuLeuAlaLysAs 1296
3721 TACCGCAACAAACCGACCTGCGCAATTCGATATGCAGAAACCTCG 3770
1296 nSerIleAsnSerAlaAsnPheValGlyAlaIleValThrAspThrSer 1313
3771 CAGCG ..... 3775
1313 IaIleIleThrAsnAlaGlnLeuAspValAlaLysAspIleGlnAlaGln 1329
3776 ..GCGCGTGGGACCTCTGTTTCGCAACCGGACCGGAAACCTTC 3822
1330 LeuGlyAsnArgLeuGlyAlaLeuArgTyLeuGly .....ThrProGln 1344
3823 GACGCGGCAATCGGCAACTCGGCAAGCGCTTCCACGCTGCTTTCGG 3872
1344 uThrAlaGlnMetAlaGly .....ProGlnAlaGlyAlaIleP 1357
3873 GCATATCGGCAATCGGCAAGTTCGACATCGCATCGCGCGCGCGGTT 3922
1357 roAlaIleValAlaIleAlaGlyAspGlnAlaValAspAsnValAlaTyGly 1373
3923 TTATGAGCGGACGCTTTCAGACG...GCATCAGAGCGCAAAATCGCGC 3969
1374 IleThrAlaLysProPheTyThrAspAlaHisGlnSerTyLysGly 1390
3970 GCGGTGTCATATACGATTCAGGCAAGATCGGCGGTT .....T 4013
1390 y .....LeuAlaGlyTyLysAlaLysThrThrGlyValAlaIleG 1404
4014 GCGCGGATTCGGCATCGACCG .....CACATCGGC 4044
1404 ILeuAspThrLeuAlaAsnAspAsnLeuMetIleGlyAlaIleAlaIleGly 1420
4045 GCACGCGCTTTCGTCCAAAAAGCGATTACGA .....TAGCA 4085
1421 IleThrLysThrAspIleLysHisGlnAspTyLysLysLysLys 1437
4086 AAGCTCAATATCGCACCCCGGCTTGATTCATCAACCGTCAACCGCGC 4134
1437 rAspValAsn .....GlyPheSerPheSerLeuTyGlyAlaG 1450
4134 ..... 4134
1450 InGlnLeuValLysAsnPhePheAlaGlnGlySerAlaIlePheSerLeu 1466
4134 ..... 4134
1467 AsnGlnValLysAsnLysSerGlnArgTyPhePheAspAlaAsnGlyAs 1483
4134 ..... 4134
1483 nMetSerLysGlnIleAlaIleGlyHisTyLysAspAsnMetThrPheGlyG 1500
4135 ..GGCATTAAGGCAAGATTATTCATCAACCGGCGGCAACATTCATC 4182
1500 LysnLeuThrValGlyTyLysAspTyLysAlaMetGlnGlyValLeuVal 1516
4183 ACGCTTATTTGAGCGCTGCTTACCGATGCGGCTTCGGCAAAAGTCCG 4232
1517 ThrProMetLacIleLeuSerTyLysLysSerSerAspLysnTyLys 1533
4233 AACGCGGTCAATACCGCGCTATTCGCGCAGATTCGCGCAA...ACC 4279
1533 sGlnThrGlyThrThrValAlaAsnLysGlnValAsnSerLysPheSerA 1550
4280 GCACTCGGAGATGGCGGTAACGCGGAAATCAAGGTTTCAACGCTGTCC 4329
1550 sPArgThrAspLeuIleValGlyAlaLysValAlaGlySerThrMetAsn 1566

```

4330 CTCACGCTGCGCCGCGCAAGGCGCAATGGACG 4368  
 1567 ILeThrAspLeuAlaValTyr...ProGluValHisAla 1578  
 seq\_name: Swissprot\_40: YDEK\_ECOLI

seq\_documentation\_block:  
 ID YDEK\_ECOLI STANDARD; PRT; 1325 AA.  
 AC P32051; P76140; P77168;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical lipoprotein ydek precursor (ORF1).  
 GN YDEK OR ORF1 OR B1510.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NC NCBI\_TaxID=562;  
 RX [1] SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Bada T., Fujita K., Hayashi K., Inada T., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.,  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [3]  
 RP SEQUENCE OF 595-1325 FROM N.A.  
 RX MEDLINE=94100243; PubMed=8274505;  
 RA Catwright P.J., Timms M.W., Litgow T., Hoef P.B., Hoogenraad N.J.,  
 RT "An Escherichia coli gene showing a potential ancestral relationship  
 to the genes for the mitochondrial import site proteins ISF42 and  
 MOM38.";  
 RL Biochim. Biophys. Acta 1153:345-347(1993).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 (Potential).  
 CC -1- SIMILARITY: TO E. COLI YFAL.  
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS  
 ISP42 AND MOM38.  
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A  
 FRAMESHIFT IN POSITION 653.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB000248; AAC74583.1; -  
 DR EMBL: D90793; BA15190.1; ALT\_INT.  
 DR EMBL: D90794; BA15197.1; ALT\_INT.  
 DR EMBL: X73295; CA51730.1; ALT\_FRAME.  
 DR PIR: S34315; S34315.  
 DR ECGene: EG11780; ydek  
 DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN. 1.

KW Hypothetical protein; Membrane; Lipoprotein; Signal;  
 KM Complete proteome.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 1325 HYPOTHEMETICAL LIPOPROTEIN YDEK.  
 FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).  
 FT CONFLICT 884 884 N -> K (IN REF. 3).  
 FT CONFLICT 1317 1317 M -> S (IN REF. 3).  
 SQ SEQUENCE 1325 AA; 136514 MW; 26A3A066FA19AD7D CRC64;  
 alignment\_scores:  
 Quality: 235.00 Length: 1487  
 Ratio: 0.338 Gaps: 83  
 Percent Similarity: 46.738 Percent Identity: 20.108  
 alignment\_block:  
 US-09-303-518d-653 x YDEK\_ECOLI ..  
 Align seg 1/1 to: YDEK\_ECOLI from: 1 to: 1325

613 CGGTGATGAGAGAGACGACCAATACCGCAAGATTCATATATGTC 662  
 25 ATGATAGlyThrSerThrValAsnLeuArglySerSerGlyLeuTh 41  
 663 AACGCGATATTCCTGGCTGCGTGGCAATACCTTGGCAAAATGAT 712  
 41 rThlySpSerArgLeuThrLeuGlyValLeuAlaLeuSerGlyS 58  
 713 CA...GCTGTGCGACAGTCACCTTAGTCGCAAAATTAACATAC 759  
 58 erAlaSerGlyAlaSerLeuGluValAspAspSerLeuThrAsn... 73  
 760 CCATATGCTTTTACCAACAGGAGGCTCATTTGGCAGACGCTCAC 809  
 74 .....IleAspThrAsp..... 77  
 810 AATGTTATATGATGACCCCAAGCAAAAGTGTATATATGCGGAT 859  
 78 ....ValAlaTyrAspAlaTyrLeuValGlyTyrGlyThrGlyVal 93  
 860 TGCACACA.....GGCAGC.....CCG 876  
 93 euAsnIleLeuAlaGlyGlyAsnAlaSerLeuThrThrIleThrSer 109  
 877 TATATGAGAAAGCAATGCGCTTCAG.....CTACTGCTAA 914  
 110 ValIleGlyAlaAsnGluAspSerGlyGlyThrValAsnValLeuGly 126  
 915 AGATTGG...TTCTATCATGAA.....AGCTTGGCG 943  
 126 yThrTPArgLeuTyrAspSerGlyAsnAlaArgProLeuAsnValG 143  
 944 GAGAT.....ACCATTCAGTATTCACGACCCACATCAAT 981  
 143 IyGlnSerGlyThrGlyThrLeuAsnIleGlyGlnGlyHisValAsp 159  
 982 GCGAATACGTTT...TTTACGACATATATATAGCGGAGGAAATCGA 1028  
 160 GlyIleGlyTyrLeuArgLeuGlySerThrGlyValGlyThrValAs 176  
 1029 TGGCAACATTAACACATATCTCTACCTTATAGTAAACACGACG 1078  
 176 nValGlyGlyGluAspSerValLeu.....Thr 186  
 1079 TTCATATGTTAATGTT.....TCTTATTCGAGACA 1110  
 186 hrGluLeuPheGluIleGlySerTyrGlyThrGlySerLeuAsnIleThr 202  
 1111 GCAGAGACCTGTTTATCATGCTGCGAGGCGGTCAACGATATGACG 1160  
 203 AspIyGlyTyrValThrSerSerIleValAlaIleLeuGlyTyrGlnAl 219  
 1161 CAGACTGATATATGAGAAATATTCCTTATTCAGCAAGAAAGGTG 1210

219 agly...Serasnlglnval...Valglnlygly...gly 232  
219 .....  
221 AATGATACCTACCGACACATCAACCAAGCGCGCGGCTTGTATTT 1260  
221 .....  
232 lurtplleuilelysasn...AspserSerileglnphe 244  
232 .....  
1261 GAG...GTAATTTTACGGTCTCGCTTAAACACGAACGATGGCAGG 1307  
1261 .....  
245 Glnlleglyasn...Glnl 250  
245 .....  
1308 GCGGCG...GTTCAATACAGTATGAGTACCTTACTTGGAA... 1350  
1308 .....  
250 ythnglyglualathrlleargllygllyleuvalthralgluhsnt 267  
250 .....  
1351 .....GTAACGGCGTGGCAACGACCGCTGTCCAAATCGCAAGCG 1395  
1351 .....  
267 hrllelleglylyasnlatthr...Glyllegly 277  
267 .....  
1396 ACCTGCTGGTTCAGCCAAAGGGAACCAAGCTCGTACGCG... 1443  
1396 .....  
278 Thrleuasnvalgln...AspGlnaspservalillethrya 290  
278 .....  
1444 .....GCGCAGGTAAAGTCATCTAGATC 1468  
1444 .....  
290 lartgargleuetyrasnlglytrpneglyasnlglythrvlaasnilesera 307  
290 .....  
1469 AGCAGCGCGATCAAGCAAGCAAAACAGCCCTTAGTGAATCGGCTG 1518  
1469 .....  
307 snasnlglyleuileasnlnlysgln...Tyrserleuvalglyval 321  
307 .....  
1519 GTCAGCGGC...AGGGGAGCGGTGCAACTGAATCCGATATCAGTTCA 1565  
1519 .....  
322 Glnaspglyserhisgllyvalvalasnvalthr... 332  
322 .....  
1566 CCGCGCAAACTCTATTTGGCTTTCGCGCGGACGTTGATTTGACG 1615  
1566 .....  
333 ...AspLysglYhistrpasmnheleugly...Thrg 343  
333 .....  
1616 GGCATTCGCTTGTCCACCGCATCAAAATACCGTGAAGGGGAGG 1665  
1616 .....  
343 lylunlathrlearglytlethrylleglyaspalaiglyaspilglnle 359  
343 .....  
1666 ATTGTCAACCAATCAAGACAAAGAAATCCAGCTTACCATTAACGCA 1715  
1666 .....  
360 Asnvalasersegllylyvalaspserglyllethralaglyne 376  
360 .....  
1716 TAAGATATTACTACACCGGCAATTAACAACATTGGATCAAAAAA 1764  
1716 .....  
376 llysglu...Thnglythrglyasnillethryvallysasplysasnserv 392  
376 .....  
1765 .....GAATTCCTACACGCTTGGCTTGGCGG 1794  
1765 .....  
392 alillethrasnleuglythrasnleuglytyraspely... 404  
392 .....  
1795 AAAGATGCAACCAAAAGACGCGGCTCAATTCATTAACACCGGA 1844  
1795 .....  
405 .....Hisgljumetasnilleaserasnln... 413  
405 .....  
1845 AGAAGCGATCGACCTTACTGCTTCCGGGAGCAAAATTTAAAC... 1890  
1845 .....  
414 .....Glyleuvalvalaserasnlglyserleuglytyrg 427  
414 .....  
1891 .....GCGCATATCAAGCAAAACGCAAACTGTTTTC 1926  
1891 .....  
427 lylulthrglyvalglyasnvalserillethrglyglymetrp... 442  
427 .....  
1927 ACGGAGACCGACCGGCTTACATCAATTAAGAAAGCGGTGCTC 1976  
1927 .....  
443 .....Glnvalasnlysasnvallythrllegly... 453  
443 .....  
1977 AAAAAAGGATATCCACAGAGAAATCGTGTGGCAACGATTTGA 2026  
1977 .....  
454 ...Valalaglyval... 457  
454 .....  
2027 TCGACCGCATTTAAAGCGGAACCTTCATATTCAGGCGGACAGCG 2076  
2027 .....  
458 .....Glyasnleuasnileserasplyglylys 467  
458 .....  
2077 GTGGTTTCCCGCAATGTCGCCAAAGTGAAGCGGATGGCATTTAAGCA 2126  
2077 .....  
468 phevalserGlnasnillethrpheleu...glyasp... 478  
468 .....  
2127 TCGACGCCAAGCATTTTCGTGTGCGACCGCATCAAGCCACATCT 2176  
2127 .....  
478 ..... 478  
2177 GTACACGTTTCGAGTGGACGGGTCTGACAAGTTGACGAAAAACATT 2226  
2177 .....  
479 ...Lysalaserglylleglythrvleuasnleumetasplathrser 493  
479 .....  
2227 ACCGACGATTAAGTATGCTTCATTTAGACCAAGCCGACATCAGAGCA 2276  
2227 .....  
494 Serpheaspthvalglylleasnvalglyasnhegly...SerGlyyl 509  
494 .....  
2277 TGTACGCTTGGCGATTCACGCTCATTTAAATCTCAGAGCTTGGCACAC 2326  
2277 .....  
509 eValasnvalaserasnlglyalathrvleuasnserthrglytyrglyphei 526  
509 .....  
2327 TCACGCGCATTTAGTCAGGCGGA...GACAGC 2358  
2327 .....  
526 leglyglyasnlaserserlyglyllevalasnileserThraspsr 542  
526 .....  
2359 CACTATACGTTACCGGACGCGACCCAC...CAAAACG 2393  
2359 .....  
543 leutrpasnleuLythrSerSerthrasnlaGlnleuGlnvalgl 559  
543 .....  
2394 C...AACCTCAGCCTGTGGGCAATGCCAAG 2422  
2394 .....  
559 yalleuglythrglyleuasnillethrglyglyllevallysa 576  
559 .....  
2423 CAACATTAAATCAAGCCCATTTAAACGGCAACATCGGCT... 2463  
2423 .....  
576 lartgaspthrGlnlleleuasnnsplyserlyGlyAspvalarg 592  
576 .....  
2464 ...TCGCAATGCT...TCATTATATCAACCAACAGC 2498  
2464 .....  
593 ValaspglyGlnasnSerleuGlnthrphesnmetyryvalglyth 609  
593 .....  
2499 CGTACAAACGCGCAGTGTGACGCTTTCGACACAGCTTAAGCAACGTA 2547  
2499 .....  
609 rserglythrglythrvleuthrvleuthrasnnglythrvleuasnvalg 626  
609 .....  
2547 ..... 2547  
626 lylglygllyvaltyrleuGlyvalphegluproalaValglythrvleu 642  
626 .....  
2548 ...ACCAT...TCGCACTCAACGCGCAATGCTC 2576  
2548 .....  
643 Asnilleglyalaiahisglyglualalaialasplaellypheieth 659  
643 .....  
2577 CCTACCGCATTAAGCAGTATTC... 2598  
2577 .....  
659 rasnlaathrvlyvalglupheglyleuglygllyvalphevalphea 676  
659 .....  
2599 ...CATTTGAAACAGC...CGCTTAC 2622  
2599 .....  
676 snhistrpasnaseraspalaiglytyrglnvalaspmetleuilethr 692  
676 .....  
2623 GGAATAATCAAGCGCGCAAGATACGCAATTAACGCAAGCGGA 2672  
2623 .....  
693 glyaspsplysaspglylys...Valillehis...Aspala 705  
693 .....  
2673 ATGACGCTG...CGTCCGCGACGGAATTAAGCAATTTAAACCTTG 2716  
2673 .....  
705 yhisThrvalpheasnlaiglyasnThrtyrserglylyThrleuvala 722  
705 .....

```

2717 ACAAGCCACCATTTACTCAATTCGGCTATCGACACGATGGCGACG 2766
      ::::: ::::: ||| ||| |||
722 snasplyleuethrillealaser.....Histhrillaasply 735
      ::::: ::::: ||| ||| |||
2767 GCGCAACCGCGGACGTGGCAGATCGCGCGCGCGCTTCGCGCGCTC 2816
      ::::: ||||| :::::
736 val...Thrglymetcliser.....Se 742
      ::::: ::::: ||| ||| |||
2817 CCTATTATCCGTTACGCGCGCAATTCGGAGATCCGTTTCAACACGC 2866
      ::::: ::::: ||| ||| |||
742 rgluvalthrillealaserPro.....GlythrL 752
      ::::: ::::: ||| ||| |||
2867 TCACGGTAAACGGCAATTGACGGTCAGGACATCCGGTTTATGCG 2916
      ::::: ::::: ||||| :::::
752 euasplileuethrillaserthrlnseralaglyasprtythrleuthrln 768
      ::::: ::::: ||||| :::::
2917 GAACCTCTTCGGCTACCGCGCGCAAAATTGAAGTGGCGAAATTCG 2966
      ::::: ||| ||| :::::
769 AlaleuLygLy.....AsplyleuethrArgValGlnleuSerSe 783
      ::::: ::::: ||||| :::::
2967 AGGCACTTACACCTTGGCTGACAAATTCGGCAACGACCGCTTAATC 3016
      ::::: ::::: ||||| :::::
783 rAsplymetrphgLythrlnsalatThrGlythrphlealagLy 800
      ::::: ::::: ||||| :::::
3017 TCGAGCAATTGACGGTAGTGAAGAAAGACACACACCGCTTCGCA 3066
      ::::: ||||| :::::
800 alAlaGlnleu.....LysasprThr..... 807
      ::::: ||||| :::::
3067 AATCTTAATTTCACCTTCGCAAAAGACACAGTGCATGCCGGCGATGCG 3116
      ::::: ||||| :::::
808 .....PheThrleuGlnArgAsp.....AsnThrAlaAlaLeuTh 819
      ::::: ::::: ||||| :::::
3117 TTAATGCTTATCCGCAAGACGGGAGTTCGGCTGCATATCCGGTCA 3166
      ::::: ::::: ||||| :::::
819 rHlsalameuLeuGlnSerAsprSerLysanThrSerValLyValG 836
      ::::: ::::: ||||| :::::
3167 AAGAACAAGAGCTTTCGCAAACTCGGCAAGGAGAAACAAGAGCC 3216
      ::::: ||||| :::::
836 LylGlnleuSerlleGlyGlyLeuAlaMetasnGlyGlyThrlelePhe 852
      ::::: ::::: ||||| :::::
3217 GCGTTCAGCGCAAAACAGGACACACTTCGCGCGCAACACAGCGGAAA 3266
      ::::: ||||| :::::
853 AsprThrAsplleProAlaAlaThrleuAla..... 862
      ::::: ||||| :::::
3267 AGCAACAGCGGCAAGCTTCAGCGCGTGCATTCGGCGGCGCAATGCCA 3316
      ::::: ::::: ||||| :::::
863 .GlnGlyTyrlleSerValAsprThrleuValGlyAlaGlyAsprTyT 879
      ::::: ::::: ||||| :::::
3317 CCGAAAGCGCAGAAAGT..... 3333
      ::::: ||||| :::::
879 hrrtyrlyserGlyArgasnTyrlGlnValasnGlyThrGlyAspralleuile 895
      ::::: ::::: ||||| :::::
3334 ...GTTCCGCAACG..... 3345
      ::::: ||||| :::::
896 AsprValProLyserProtyrAsnAsprMetAlaAsnAsprleuThrTh 912
      ::::: ::::: ||||| :::::
3345 ..... 3345
      ::::: ||||| :::::
912 rLeuAsnleuLeuGlnHisAsprSerHsValGlyValGlnleuValL 929
      ::::: ::::: ||||| :::::
3346 ...GCCGCGACGAGCGGCGGAAATCGCGCATTTATGACGCGGAGAA 3393
      ::::: ::::: ||||| :::::
929 ysAlaGlnThrVallleGlySerGlyGlySerleuThrleuArgAsprleu 945
      ::::: ::::: ||||| :::::
3394 GAGAAAAACGCGTCAGCGCGATAAAGACACCGCGCTTCGCAACAGCG 3443
      ::::: ::::: ||||| :::::
946 GlnGlyAsprGlyValGlnAlaAsprTyThrleuHislleAlaGlnasnG 962
      ::::: ::::: ||||| :::::
3444 CGAAGCGGAAACCGG.....CCGCGTACACCGCGCTTCG 3478
      ::::: ::::: ||||| :::::
962 yThrValValAlaGlnGlyAsprTyGlyrphArgleuThrThrAla.... 977

```

```

3479 CCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACGCGACCGCCAAACG 3528
      ::::: ::::: ||||| :::::
977 ..... 977
      ::::: ||||| :::::
3529 CAACCCCAACCGACGCGACCTGATACCGCTTATGCCAATAGCGGTTT 3578
      ::::: ::::: ||||| :::::
978 .....ProGlyasnGlyLeu.....TyValAsnTyGlyLe 988
      ::::: ::::: ||||| :::::
3579 G..... 3579
      ::::: ||||| :::::
988 ulysAlaLeuAsnlleHisGlyGlyGlnLySleuThrleuAlaGlnHisG 1005
      ::::: ::::: ||||| :::::
3580 ...AGTGAATTTCCGCGCACGCTCAGACGCGTTTCCGCGTACAGAGAA 3627
      ::::: ::::: ||||| :::::
1005 LylGlyAlaTyrlGlyAlaThrAlaAsprSerAlaLyslleGlyGlyGln 1021
      ::::: ::::: ||||| :::::
3628 TTGACCGCGCTGTTTCGCGAAGACCGCGCGACCGCGTTTGCAAGCG 3677
      ::::: ::::: ||||| :::::
1022 GlyAsprleuAlaLeuAsnThrValArgGlnValSerleu...SerasnG 1037
      ::::: ::::: ||||| :::::
3678 CATCCGGGACACCAACACTACCGCTTCGCAAGATTTCCGCGCTTACCGC 3727
      ::::: ::::: ||||| :::::
1037 yGlnAsnAsp.....TyrlGlnGly.....AlaThrTyValG 1048
      ::::: ::::: ||||| :::::
3728 AACAAACCGACCTCGCG.....CAATCGGTATGCAAGAA 3762
      ::::: ||||| :::::
1048 lmetGlyThrleuArgThrAsprAlaAsprLyAlaLeuGlnAsnThrArg 1064
      ::::: ||||| :::::
3763 AACCTCGGCGAGCGCGCGCTCGCGATC.....CTGTTTCGCAACACG 3806
      ::::: ::::: ||||| :::::
1065 GlnleuAsnlleSerAsnAlaAlaIleValAsprleuAsnLysThrGln 1081
      ::::: ::::: ||||| :::::
3807 GACCGGAAACACTTCGACGACGCGATCGCAACTCGCGAGGCTTCCCG 3856
      ::::: ||||| :::::
1081 nThrValGlnThrPheThrGlyGlnMetGlySerThrValleuPheLyS 1098
      ::::: ::::: ||||| :::::
3857 AGCGTCCGTTTCGCGCAATACGCGATCGCGAGTTTCGACATCGGCATC 3906
      ::::: ||||| :::::
1098 lGlnLyAlaLeuThrValAsnLySly..... 1106
      ::::: ||||| :::::
3907 AGCGCGGCGCGGCTTTAGTAGCGAGCGCTTTCAGACGCGATCAGAGG 3956
      ::::: ||||| :::::
1107 .....GlylleSerGlnGlyGlnleuThrGlyGlyGlyAsnle 1119
      ::::: ||||| :::::
3957 CAAATCCGCGCGCGCGCTGCTGATTCAC...GGCATTCAGCAAGATACC 4003
      ::::: ::::: ||||| :::::
1119 uasnValThrGlyGlyThrleuAlaIleGlnGlyLeuAsnAlaArgTyra 1136
      ::::: ::::: ||||| :::::
4004 GCGCAGGTTTCGCGGATTCGCGATCGACCGCGCACATCG..... 4042
      ::::: ||||| :::::
1136 snAlaLeuThrSerlleSerProAsnAlaGlnValSerleuAsnThr 1152
      ::::: ||||| :::::
4043 .....GCGCAACGCGCTATTTTCGCAAAA 4067
      ::::: ||||| :::::
1152 rGlnGlyLeuGlyArgGlyAsnlleAlaAsnAsprGlyleuThrleuL 1169
      ::::: ::::: ||||| :::::
4068 AGCGGATTCAC...GATPAGAAACGTCATATGTCACCGCGGCGCTTG 4114
      ::::: ::::: ||||| :::::
1169 ysAsnValThrGlyGlyleuArgAsnSerlleSerGlyLyGlylleVal 1185
      ::::: ||||| :::::
4115 CATTCACCGCT.....ACCGCGCGGCG 4137
      ::::: ||||| :::::
1186 SerAlaThrAlaArgThrAsprValGlyleuAsprGlyAsnAsnSerAla 1202
      ::::: ||||| :::::
4138 ATTAAGCGAGATTTATTCATCAACCGCGCAACACATTTCCATCAGCGC 4187
      ::::: ||||| :::::
1202 evalGlyGlnPheAsnlleAsprThrGlySerAlaLeuSerValasnG 1218
      ::::: ||||| :::::
4188 TTATTGAGCGCTGCTTACGAGTTCGCGCTTCGCGAAATTCGCAACG 4237
      ::::: ||||| :::::
1218 uGlnLyAsnleuGly.....AsprAlaSerVal..... 1227
      ::::: ||||| :::::
4238 GCGTCATATCCGCGCTATTTGCGCAGCATTTTCGCAAAACCGCGATGCG 4287

```











```

512 aglyThrIleThrLeuAspGlySerAlaThrIleThrGlyAspIleGly 529
    ||| |||||:|||||: |||: |||:
1481 ATCAAGGCAAAAAAAGCCTTATGGAATCGCGCTTGCGTGCAC...GGC 1527
    :: ||| |||||:|||||: |||: |||:
529 snlaGlyGlyAlaAlaLeuGlnArgIleThrLeuAlaAsnAspAla 545
    :: ||| |||||:|||||: |||: |||:
1528 AGGGGAGCGGTGCACCTGAATGCCGATTAATCACTTCAACCCGACAAACT 1577
    :: |||: |||: |||: |||
546 LysLysThrLeuThrLeuGlyGlyAlaAsn..... 555
    |||: |||: |||: |||
1578 CTTTTCGCTTTCGCGCGGAGCTTGATTTG.....AACGGGCATT 1621
    :: |||: |||: |||: |||
556 .IleIleGlyAlaGlyGlyThrIleAsnLeuGlnAlaAsnGlyGlyT 572
    :: |||: |||: |||: |||
1622 CGCTTCGTTCCACCGCATTCAAAT..... 1647
    :: |||: |||: |||: |||
572 hrLysLysLeuThrSerThrGlnAsnAlaIleValAlaAspHeu 588
    :: |||: |||: |||: |||
1648 .....ACCGATGAGGGCGGATGATGTCAACCAATCAAGACAA 1688
    |||||: |||: |||: |||
589 AlaIleAlaThrAspGlnThrGlyAlaAspAlaSerSerLeuThrAs 605
    :: |||: |||: |||: |||
1689 AGAATCCACCGTTTACCATTCAGAGCAATGAATATTAATTAACCGGCA 1738
    : ::|||: |||: |||: |||
605 nlaGlnThrLeuThrIleAsnGlyLys.....IleGlyThrIleGlyA 620
    |||||
1739 ATACACAACTGGATAGCAAAAAAATTCCTACACGTTGGTTT 1788
    |||||
620 laAsnAsn..... 622
    |||||
1789 GCGGAGAAAGATGCAACCAAAACGAGCGGCTCAATCTGAATTACCA 1838
    ||||| |||||: |||: |||: |||
623 .....LysThrLeuGlyGlnPheAsnIleGly..... 631
    |||||: |||: |||: |||
1839 ACCGGAAGAGCGATCGACTTACTGCTTCCGCGGGAACAATTTAA 1888
    ::|||: |||: |||: |||
632 .....SerSerLysThrValLeuSerAsnGlyAsnValAlaIleA 645
    |||||
1889 AC.....GCGAAT..... 1896
    |||||
645 snGluLeuValIleGlyAsnAspGlyAlaValGlnPheAlaHisAspThr 661
    |||||
1897 .....ATCAGCGCAACA.....AACGCAAACTGTTT 1925
    |||||: |||: |||: |||
662 TyrLeuIleThrArgThrAsnAlaIleGlyGlnGlyLysIleIlePh 678
    ::|||: |||: |||: |||
1926 C.....ACGGCAGACCGACACCGCCCTACCAATCATTTAG 1963
    ::|||: |||: |||: |||
678 eAsnProValValAsnAsnGlyThrThrLeuAlaIleGlyThrAsnLeuG 695
    |||||
1964 GAAGGGGGTGTCAAAAATGGAAGTATCCCAAGAGGAATCGTGTG 2013
    |||||: |||: |||: |||
695 LysSerAlaThrAsnProLeuAlaGluIleAsnPheGlySerLysGlyVal 711
    ::|||: |||: |||: |||
2014 GACAACGATTCGATCGACCGCATTTAAAGCGAAAACTCCATTT.. 2061
    :: |||: |||: |||: |||
712 AsnValAspThrValLeuAsnValGlyGlyValAlaAsnLeuTyrAlaThr 728
    :: |||: |||: |||: |||
2062 .....CAGG 2065
    |||||
728 rAsnIleThrThrAspAlaAsnValGlySerPheValPheAsnIleG 745
    |||||
2066 GCGGACAAGCGGTGTTCCGCAATGTGCGCAAGTGAAGCGGATTTGG 2115
    |||||: |||: |||: |||
745 LysGlyThrAsnIleValSerGlyThrValGlyGlyGlnGly..... 759
    ::|||: |||: |||: |||
2116 CATTTAAGCATACGCCCAAGCAGTTTTCGTGTCACCGCATCAAG 2165
    ::|||: |||: |||: |||
760 .....AsnLysPh 762
    |||||
2166 CCAACAATCTGTACACGTTGCGACGTGACGGGTGTGACAAGTTGTACCG 2215
    ::|||: |||: |||: |||

```

---

```

762 eAsnThrValAlaLeuGluAsnGlyThrThr..... 772
    ::|||: |||: |||: |||
2216 AAAAAACATTACCAGCATTAAGTATGCTTCATTGACGAAGCCGAC 2265
    ::|||: |||: |||: |||
773 .....ValLysPheLeuGlyAsnAlaThr 780
    ::|||: |||: |||: |||
2266 ATCAAGCAATGTGACGCTTCCGATTCAGCTCATTTAAATCTACAGG 2315
    ::|||: |||: |||: |||
781 PheAsnGlyAsnThrThrIleAlaAlaAsnSerThrLeuGlnIleGly 797
    ::|||: |||: |||: |||
2316 ACTTCCACACTCAACGCGCATCTTAAGCGGGGAGACCGGACACTATA 2365
    ::|||: |||: |||: |||
797 YasnThrThrAlaAspCysValAlaSerAlaAspGly.....T 810
    ::|||: |||: |||: |||
2366 CGGTTCAGCGACGCGCCACCCAAAACGCAACTCAGCTCTGTGGCAAT 2415
    ::|||: |||: |||: |||
810 hrGlyIleValAluPheValAsnThrGlyProIleThrValThrLeuAsn 826
    ::|||: |||: |||: |||
2416 GCCCAAGCA.....ACATTTATCAACCCACATTTAAACGCCAA 2453
    |||||: |||: |||: |||
827 LysGlnAlaAlaProValAsnAlaLeuLysGlnIleThrValSerGlyPr 843
    ::|||: |||: |||: |||
2454 CACATCGGCTTCGGACAATGCTTCAATTTAATCTAAGCAAC..... 2496
    ::|||: |||: |||: |||
843 OGlyAsnValValIleAsnGluIleGlyAsnAlaGlyAsnHisHisGlyA 860
    ::|||: |||: |||: |||
2497 .....GCGTACAAAACGCGACTGTG..... 2517
    ::|||: |||: |||: |||
860 laValThrAspThrIleAlaPheGlnAsnSerSerLeuGlyAlaValAl 876
    ::|||: |||: |||: |||
2518 .....ACGCTTCGCA 2528
    ::|||: |||: |||: |||
877 PheLeuProArgGlyIleProPheAsnAspAlaGlyAsnThrMetProle 893
    ::|||: |||: |||: |||
2529 CAACGCTAAGGCAAAACGTAAGCCATTCCGCACTCAACGCG..... 2568
    ::|||: |||: |||: |||
893 uThrIleLysSerThrValGlyAsnLysThrAlaLysGlyPheAspValP 910
    ::|||: |||: |||: |||
2569 .....AATGTCCTAGCCGCTAAGGCAGTA 2595
    ::|||: |||: |||: |||
910 roSerValValValLeuGlyValAspSerValIleAlaAspGlyGlnVal 926
    ::|||: |||: |||: |||
2596 TTCATTTTGAAGAAAGCGCTTACCGGAAATACGCGC..... 2637
    ::|||: |||: |||: |||
927 IleGlyAspGlnAsnAsn.....IleValGlyLeuGlyLe 938
    ::|||: |||: |||: |||
2638 .GCGACGATACGCGATTACACTTAAAGACAGCGAATGACGCTCCGT 2686
    ::|||: |||: |||: |||
938 uGlySerAspAsnGlyIleIleValAlaSerAlaThr.....ThrLeuTyrA 953
    ::|||: |||: |||: |||
2687 CGGGCAGCGAATTAAGCAATTTAAACCTTGACACACGCCACATTAACATC 2736
    ::|||: |||: |||: |||
953 laGly.....IleSerThrLeuAsnAsnAsnGlnGlyThrValThrLeu 967
    ::|||: |||: |||: |||
2737 AATTCGCGCTATGACACGATGCGGCA.....GGCGCGCAACCGG 2777
    ::|||: |||: |||: |||
968 SerGlyGlyValProAsnThrProGlyThrValTyrGlyLeuGlyThrAl 984
    ::|||: |||: |||: |||
2778 CAGTGGCGGAGATGCGCGCGCGCGCTTCGCGCGCTTCCTATTATTCGG 2827
    ::|||: |||: |||: |||
984 yIleGlyAlaSer.....LysPheLysGlnV 993
    ::|||: |||: |||: |||
2828 TTAAGCGCGCACTTCGCGAGATCCGTTTCACACACGCGTGAGGTAAC 2877
    ::|||: |||: |||: |||
993 alThrThrThrAspTyrAsnAsnLeuGlyAsnIleIleAlaThrAsn 1009
    ::|||: |||: |||: |||
2877 ..... 2877
    ::|||: |||: |||: |||
1010 AlaThrIleAsnAspGlyValThrValThrThrGlyGlyIleAlaGlyI 1026
    ::|||: |||: |||: |||
2878 .....GCCAAATTGAACGCTGAGGGAACAT 2902
    ::|||: |||: |||: |||
1026 eGlyPheAspGlyLysIleThrLeuGlySerValAsnGlyAsnGlyAsnV 1043
    ::|||: |||: |||: |||

```











```

3569 ATAGCGTTGAGTGAATTTCCGCCAGCGCTCAAGCGCTTTCCGCCGTA 3618
      ::::: ||||| ||||| :::::
3327 ValGluThrIaGlnProLeuProIaGlnProValaIa..... 3339
3619 CAGAGCAATTTGACCGCGCTTTGGCAGACCGCGCGACCGCGTTG 3668
      ::::: ||||| |||||
3340 .....AlaGlnValaIaProValaThr..... 3346
3669 GACAGCGCATCCGGACACCAACACTACCTTCGCAAGATT..... 3712
      ||||| ::::: |||||
3347 .....ProIaGlnValaIaGlnValaIaGlnValaIa 3357
3713 .....TCGCGCGCTACCGCGCCACACACCGAC 3738
      ::::: ||||| :::::
3358 ValProArgProIaGlnValaIaGlnProLeuProIaGlnProVala 3374
3739 CTCGCGCAATCGCTATGCAAAAAACCTCGCAGCGCGCGCTCGCAT 3788
      ::::: ||||| :::::
3374 ValaIaGlnValaIa.....ThrThrProIaValaIaGlnProIaGln 3389
3789 CCTGTTTTCGACACCGACCGCAACACTTCGACGACG..... 3829
      ::::: ||||| :::::
3389 IalysValaIaGlnThrValaIaGlnProValaIaGlnThrIaGlnPro 3405
3830 .....GCATCGCGCACTCGCAGCGCGCTTCGCCACGCT 3861
      ||||| ::::: |||||
3406 ProIaGlnProLeuProValaIaGlnValaIaGlnProIaGlnProVala 3422
3862 G.....CCGTTTTCGCGCAATTCGCGCGCGCGCAG 3890
      ::::: ||||| :::::
3422 IeGluThrIaGlnProLeuProIaGlnProIaGlnProIaGlnProIa 3439
3891 GTTCGACATCGCATCGCGCGCGGCTTTAGTAGCGCGACCGCTT 3940
      ||||| ::::: |||||
3439 IyProValaIaGlnValaIaGlnValaIaThrVala..... 3449
3941 CAGAGCGCATCAGAGCAAAATCCCGCGCGCTCGCATTCAGCGCAT 3990
      ||||| ::::: |||||
3450 ThrThrValaIaGlnValaIaGlnSerIaIaProIaGlnProIaGln 3462
3991 CAGGCAAGTACCGCGAGGTTTCGCGGATTCGCGCATCGCAACCGCAT 4040
3462 ..... 3462
4041 CGCGCGCAACCGCTATTCGTCCAAAAACGCGATTACCGATACGAAACG 4090
      ||||| ::::: |||||
3463 .....ProValaIaIaGln 3467
4091 TCATATATCGCACCGCGCGCTTCGATTCACACCGTACCGCGCGCAT 4140
      ||||| ::::: |||||
3467 nProIaGlnProIaGlnProIaGlnProIaGlnProIaGlnProIaGln 3483
4141 AAGCGAGATTATTCATCAACCGCGCGCAACACATTCATCAGCGCTTA 4190
      ::::: ||||| :::::
3484 ArgPro.....LysProGlnLys..... 3489
4191 TTGAGCGCTGCTCATACCGATCCCGCTTCGCGCAACGCGCGCG 4240
      ||||| ::::: |||||
3490 .....ThrThrProLeuSerGlyArgHisValaIaGln 3501
4241 TCATACCGCGCGCTATTCG.....GGCGAGATTTCGCGCAACCGCGC 4281
      ::::: ||||| :::::
3501 IuGlnValaIaGlnValaIaGlnValaIaGlnValaIaGlnValaIaGln 3517
4282 AGTGGCGAGATGGCG..... 4296
      ||||| ::::: |||||
3518 SerLeuProGlnGlyLysLeuProIaGlnProIaGlnProIaGlnProIa 3534
4297 .GTAAAGCGCGCAATCAAGGTTTCAGCGCTTCCTCCAGCGCTCCCGC 4345
      ||||| ::::: |||||
3534 pGluAsnGlyLysProGlnThrIaGlnValaIaGlnValaIaGlnVala 3551

```

```

seq_name: SwissProt_40:HX2_HAEMIN
seq_documentation_block:
ID HX2_HAEMIN STANDARD; PRT; 928 AA.
AC P45354;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
DE protein A).
GN HXUA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-35.
RX STRAIN=DL42 / SEROTYPE B;
RX MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.;
RT "The 100 kDa heme:hemopexin-binding protein of Haemophilus
RT influenzae: structure and localization."
RL Mol. Microbiol. 13:863-873(1994).
RN [2]
RP SEQUENCE OF 1-30 FROM N.A.
RX STRAIN=DL42 / SEROTYPE B;
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yogev R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
RT heme:hemopexin by Haemophilus influenzae type b."
RL J. Bacteriol. 177:2644-2653(1995).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC DR EMBL: U08348; AAA74138.1; -
CC KW TRANSPORT; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 928 HEME/HEMOPEXIN-BINDING PROTEIN.
FT DOMAIN 101 679 6 x 6 AA APPROXIMATE REPEATS.
FT REPEAT 101 106 1-1.
FT REPEAT 205 210 1-2.
FT REPEAT 279 284 1-3.
FT REPEAT 410 415 1-4.
FT REPEAT 635 640 1-5.
FT REPEAT 674 679 1-6.
FT DOMAIN 149 172 4 x 6 AA APPROXIMATE TANDEN REPEATS.
FT REPEAT 149 154 2-1.
FT REPEAT 155 160 2-2.
FT REPEAT 161 166 2-3.
FT REPEAT 167 172 2-4.
SQ SEQUENCE 928 AA; 101228 MW; 67D45466A4B92390 CRC64;

```

alignment\_scores:      Quality: 215.50      Length: 1142  
Ratio: 0.457                              Gaps: 57

Percent Similarity: 41.331 Percent Identity: 19.527

alignment\_block:

US-09-303-518D-653 x HXA2\_HAEIN ..

Align seg 1/1 to: HXA2\_HAEIN from: 1 to: 928

```

451 AACGGCCATCTTATGGCGCGATTCATATGCCCGTTGCAAAATT 500
    ::::| | | | | | | | | | | | | | | | | | | | | |
16 SerGlyAlaIleValSerThrProAspPheProGlnHisIleSth 32
    ::::| | | | | | | | | | | | | | | | | | | | | |
501 TGTACAGATGACGAACCTGTGATGACCACTTATGATGAGGTGGA 550
    ::::| | | | | | | | | | | | | | | | | | | | | |
32 rValPheGlyThrValThrIleGluSthThr 43
    ::::| | | | | | | | | | | | | | | | | | | | | |
551 AATAAGCTGAT.....TTAAATAAATACCTGATGCTGTGCAATC 591
    ::::| | | | | | | | | | | | | | | | | | | | | |
44 .....AlaAspLysMetThrIleGlyGlnLysSerAspLysAlaGlnIle 58
    ::::| | | | | | | | | | | | | | | | | | | | | |
592 GGAGCAGCAGACAAATTTGGCGGTCTGATGA..... 624
    ::::| | | | | | | | | | | | | | | | | | | | | |
59 Asp.....TrrLysSerPheAspLleGlyGlnLysG1 70
    ::::| | | | | | | | | | | | | | | | | | | | | |
625 .....GACGAACCAATTAACCGCAAGTTTCATATCATATTGCAA 664
    ::::| | | | | | | | | | | | | | | | | | | | | |
70 uValLysPheGlnIleProAsnGlnHis.....AlaV 81
    ::::| | | | | | | | | | | | | | | | | | | | | |
665 GCGCATATTTCTGGCTGCTGCGTGCAATACCTTGCACAAAATGCA... 711
    ::::| | | | | | | | | | | | | | | | | | | | | |
81 alaIaIrrAsnArgValIleGlyGlnLysAlaSerGlnIleGlnLys 97
    ::::| | | | | | | | | | | | | | | | | | | | | |
712 ...TCAGTGTGGCAGCATCACTAGTAGAGCAAAATAATTAACTAG 758
    ::::| | | | | | | | | | | | | | | | | | | | | |
98 LeuThrAlaAsnGlyLysValTyrLeuAla.....As 108
    ::::| | | | | | | | | | | | | | | | | | | | | |
759 CCCATATGCTTTTACCAACAGAGGCTCATTTGGCAGACGTGGCTCAC 808
    ::::| | | | | | | | | | | | | | | | | | | | | |
108 nProAsnGlyValIleIleThrGlnGlyAla..... 118
    ::::| | | | | | | | | | | | | | | | | | | | | |
809 CAATGTTATCTATGATGCCCAAAAGCAAAAGTGTTAATTAATGGGTA 858
    ::::| | | | | | | | | | | | | | | | | | | | | |
119 .....GluIleAsnValAlaGlyLeu 125
    ::::| | | | | | | | | | | | | | | | | | | | | |
859 TTGCAACAGCAGCAACCCCTATATAGAAAGCAATGGCTCCAGTACT 908
    ::::| | | | | | | | | | | | | | | | | | | | | |
126 LeuAlaThrThrLys.....AspLeuG1 133
    ::::| | | | | | | | | | | | | | | | | | | | | |
909 TCGTAAAGATTGGTTCTATGATGAATCTTGTGAGATACCCATTGAG 958
    ::::| | | | | | | | | | | | | | | | | | | | | |
133 uArg..... 134
    ::::| | | | | | | | | | | | | | | | | | | | | |
959 TATTCACGAACCATCAAAATGGGAATACTTTTAAACGACAAATAT 1008
    ::::| | | | | | | | | | | | | | | | | | | | | |
135 .....IleSerGlnAsnSer 139
    ::::| | | | | | | | | | | | | | | | | | | | | |
1009 AATGCGCAGAGAAAATGATGCCAAACATTAACATATCTCTCACTTA 1058
    ::::| | | | | | | | | | | | | | | | | | | | | |
140 AsnSer.....TyrGlnPheThrAr 146
    ::::| | | | | | | | | | | | | | | | | | | | | |
1059 TAGATTAAAAACAGACACCGTTCAATTTTAAATGTTTCTTATCCAGA 1108
    ::::| | | | | | | | | | | | | | | | | | | | | |
146 gArgThrLysAspArgGlnVal..... 153
    ::::| | | | | | | | | | | | | | | | | | | | | |
1109 CAGCAAGAACCTGTTATCATGCTGCAAGTGGGCTCAACAGATTATGCA 1158
    ::::| | | | | | | | | | | | | | | | | | | | | |
154 ..LeuLysGlnGlyLeuValIleuLysAspGlyGlnValValLysGlnGly 169
    ::::| | | | | | | | | | | | | | | | | | | | | |
1159 CCCAGACTGAATATGAGAAAATATTTCC.....TTTATGACAA 1199
    ::::| | | | | | | | | | | | | | | | | | | | | |
170 GlnValIleAsnGlnGly...AsnIleThrAlaGlnAspPheValIle 185
    ::::| | | | | | | | | | | | | | | | | | | | | |
1200 AGGAAGAAGTGAATGATCTTACAGACATCATCCAGGCGGGCG 1249
    ::::| | | | | | | | | | | | | | | | | | | | | |

```

```

185 uAsnGlyAspClnValIleAsnLysGlyAsnIleAsn..... 197
    ::::| | | | | | | | | | | | | | | | | | | | | |
1250 GTTTCATATTTGAGGCGATATTTTACGGCTCGCCCTAAAAACAGCAAGC 1299
    ::::| | | | | | | | | | | | | | | | | | | | | |
198 .....ValGlnLysAsnSerThrIleAsnGlnLys..... 207
    ::::| | | | | | | | | | | | | | | | | | | | | |
1300 TGGCAAGCGCGCGCTTCATATCATGATGAGTGGCAGTACCTTACTTGAA 1349
    ::::| | | | | | | | | | | | | | | | | | | | | |
208 .....ValTyrLeuSerSerGlyIrrAsnPheThr 219
    ::::| | | | | | | | | | | | | | | | | | | | | |
1350 ACTA.....AACGCGTGGCAACAGACCGCTGTCCAAATCGGCAAG 1393
    ::::| | | | | | | | | | | | | | | | | | | | | |
219 rLeuProAspSerGlyIleSer.....ValAlaLeuG 230
    ::::| | | | | | | | | | | | | | | | | | | | | |
1394 GCACCGCTGCTGTCACGCAAGAGGCAAAACAGAGCTCGTCAAGCGTG 1443
    ::::| | | | | | | | | | | | | | | | | | | | | |
230 uAspAsnThrValIleGlnGlyIleValLysAsnGlnGlySerIleLysAla 246
    ::::| | | | | | | | | | | | | | | | | | | | | |
1444 GCGCAGCGTAAAGTCATCTTAGATCAGCAGCGGACGATCAAGCAAAA 1493
    ::::| | | | | | | | | | | | | | | | | | | | | |
247 GlyGln.....IleThrLeuSerAla.....LysGlyArgLys 257
    ::::| | | | | | | | | | | | | | | | | | | | | |
1494 ACAAGCTTACTGAATCGCTTGTGTCAGCGC..... 1527
    ::::| | | | | | | | | | | | | | | | | | | | | |
257 sGlnAlaLeuAspSerLeuValMetAsnAsnGlyValLeuGlnAlaThrL 274
    ::::| | | | | | | | | | | | | | | | | | | | | |
1528 .....AGGGAGCGGTGCAACTGAATGCGCATAT...CAG 1560
    ::::| | | | | | | | | | | | | | | | | | | | | |
274 yValSerAsnLysAsnGlyLysValValLeuSerAlaAspAsnValGln 290
    ::::| | | | | | | | | | | | | | | | | | | | | |
1561 TTCAACCCCGCAAACTATTTGCGCTTTCGCGGCGGACGTTTGATTT 1610
    ::::| | | | | | | | | | | | | | | | | | | | | |
291 LeuAsnAsnGlnSer.....AsnI1 297
    ::::| | | | | | | | | | | | | | | | | | | | | |
1611 GAACGGCGCTTGGCTTGTTCACCGCATTCAAATACCGATGAAGGG 1660
    ::::| | | | | | | | | | | | | | | | | | | | | |
297 eLysGlyGlnIleValThrPhe.....GlyA 306
    ::::| | | | | | | | | | | | | | | | | | | | | |
1661 CGATGATGTCAACACATCAAGCAAGAAATCCACGCTTACATTACA 1710
    ::::| | | | | | | | | | | | | | | | | | | | | |
306 lAspValThrSer..AsnLysGlnLeuLysAspAsnIleLysThr 321
    ::::| | | | | | | | | | | | | | | | | | | | | |
1711 GGCATTAAGATATTACTCAACCGGCAATATCAACACTTGATAGCA 1760
    ::::| | | | | | | | | | | | | | | | | | | | | |
322 SerLysThrGlySerLysValThrSerProLysIleAsnPhe...ThrG1 337
    ::::| | | | | | | | | | | | | | | | | | | | | |
1761 AAAAGAAATTCCTCAACAGCTTGTGTCGAGAAAGATGCAAC... 1806
    ::::| | | | | | | | | | | | | | | | | | | | | |
337 yLysSerValAsnIleAsnGlyAsnPheGlyArgGlnAspSerThrThrH 354
    ::::| | | | | | | | | | | | | | | | | | | | | |
1807 .....AAACGACAGCGGCGCTCAATCTGAATTAC 1836
    ::::| | | | | | | | | | | | | | | | | | | | | |
354 lAspValThrSerPheLysLysLeuAsnThrGlnValAsnIleAspVal 370
    ::::| | | | | | | | | | | | | | | | | | | | | |
1837 CAACCGGAGAA.....GGGATCGGACT... 1860
    ::::| | | | | | | | | | | | | | | | | | | | | |
371 ProAspAsnGlnAsnIleArgIleAlaAspIleGlnuAspAsnThrGlyTh 387
    ::::| | | | | | | | | | | | | | | | | | | | | |
1860 ..... 1860
    ::::| | | | | | | | | | | | | | | | | | | | | |
387 rGlyThrThrGlyThrGlyThrSerSerPheIleGlnThrGlyAlaLeuS 404
    ::::| | | | | | | | | | | | | | | | | | | | | |
1861 .....TTACTGCTTTCCGCGGAGAACAAATTTAAACGGCAT... 1896
    ::::| | | | | | | | | | | | | | | | | | | | | |
404 eSerLeuLeuAlaAsnAsnGlyLysValAsnLeuLysGlyAsnVal 420
    ::::| | | | | | | | | | | | | | | | | | | | | |
1896 ..... 1896
    ::::| | | | | | | | | | | | | | | | | | | | | |
421 AsnIleSerGlyArgIleHisIleAspSerPheArgGlySerAspSerle 437
    ::::| | | | | | | | | | | | | | | | | | | | | |
1897 .ATACGCAACAAC..... 1911
    ::::| | | | | | | | | | | | | | | | | | | | | |
437 uLeuLysLeuThrAsnLysGlyHisIleAspIleAsnAlaAspIleH 454
    ::::| | | | | | | | | | | | | | | | | | | | | |

```

```
1912 .....GGCAACTGTTTTC..... 1926
1913 .....|||||
454 isserlysglyargleuphepheiethrserleuclnslngluasp 470
1927 .....AGCGGACGACGACCGCAGCCCTACATCATTTAGGAAGCG 1970
1928 .....|||||
471 pheylserasnlethrlethrapsrlyslleasnleuglyasncl 487
1971 G.....T 1972
487 yalameglyleuglyargservalasplulysasptyraspasnargt 504
1973 GGTCAAAAATGGAAGGTATCCACAAGAA.....CAAACTGTGGGAG 2016
1974 .....|||||
504 rpolnlystrngllyserglinarglyslpheasplvalylsmtser 520
2017 AAC.....GATTGATCGACCGCACATTTAAAGC..... 2046
2018 .....|||||
521 asnlvalglupheasnlevalaspsvalilleuallaglylphcgl 537
2047 .....GAAACTGCCATA 2059
537 ulysvalasnleasplysllevalalathrglylnthrapsnphetyr 554
2060 TTCAGGCGGACAAAGCGGTGTTTCCCGCAATGTGCCAAAGTGGAA... 2106
2061 .....|||||
554 leaspllygly.....ValserArgasnlyarglystrcglulyr 567
2107 GGCAGTTGGCATTTAAGCAATGACGCCCAAGCA..... 2139
2108 .....|||||
568 glyvalleuaspluaspslysrthrnglneusergluleuasngl 584
2140 .....GTTTGGGTGTCGACCGCATCAAGCCACA 2170
584 yargargargtrpglytrtyrtyraspleuclneuasphetasnarg 601
2171 CAATCTGACAGTTCGACCTGACGGGTGCACAACTTTGACGAAAAA 2220
2172 .....|||||
601 latyrlleuylrthrphaspleuaphealathrlyasnthrlyargser 617
2221 ACCATTACCGAC..... 2232
2222 .....|||||
618 thrilleysasptthrngluleasnleaserasnserasnilleasnleu 634
2233 .....GATTAAGTATGTTGCTTCATTGAGCA 2257
2234 .....|||||
634 sasnglyphevalhlisleuallaglyllyslleuaspsasnserl 651
2258 AGACCGACATC.....AGAGCAATGTACCGCTTGGCGAT 2292
2259 .....|||||
651 ysileaspllethrphaspslyaspsasnsergluaspthrleuallagln 667
2293 CACGCTCATTTAAATCTCAGAGA...CTTGCCACATCAACGCGCAATCT 2339
2294 .....|||||
668 thrasnargleuglymetasnlyllyslvalsermetlleasnserhl 684
2340 TAGTCAGAGCGGACAGACGACTTACGCTTACGCGCAACGCCACAAA 2389
2341 .....|||||
684 elysllevalglyasplulysgllylleserprothrnglythrtyra 701
2390 ACGCAACCTC.....AGCCTGCGGCGCATGCCCAAGCAACATTT 2430
2391 .....|||||
701 latnmepheleuileglyleuileuglyllyslserserlelph 717
2431 AATCAAGC.....ACATTAACGCAACACACATCGGC 2462
2432 .....|||||
718 vallysserhlslnglyltyrthrphelystrharspelyasnthrlysl 734
2463 TTCGACAAATGCTCA.....TTTAAATTA 2488
734 ealeglyllystrlyserlysgluaspleuyllethrallaleasnthr 751
```

```
2489 GCACAAACCCGCTACAA.....AACGCAGCTTGACCGCTTTC 2526
2490 .....|||||
751 lyglyarglalaalaglulvalleuileasnlylaleuglyserala 767
2527 GACACGCTTAAGCAACGATTCGCGCATTCGCAACGCGCATTCGCTC 2576
2528 .....|||||
768 aspsan...AspAlaasnlelelaasmetalaiphetrilleglyaspse 783
2577 CCTAGCCGATTAAGCGAGTATTCATTGTTGAAACAGCGCTTACCGGAA 2626
2578 .....|||||
783 ralaasnthrlystrthr.....llegluasnlalaspelthrlel 798
2627 AATTCAGCGGCGCAAGATACGCGATTACCTTAAAGACAGCGAATGG 2676
2628 .....|||||
798 euallaproasnlyglythralthrleuserserlyaspsvalglu... 813
2677 ACGCTGCCGTGGGACAGGAAATAGCAATTTAACTTGACAAACGCCAC 2726
814 .....llegl 815
2727 CATTAACCTCAATTCGCGCTATCGACAGATCGCGAGCGCGCAACCG 2776
2728 .....|||||
815 uvallysprowasnserasnpe..... 822
2777 GCAGCGCGGACAGATGCGCGCGCGCGCTTCCGCGCTTCCATTATTC 2826
823 .....Thrphiegluleuproarg..... 829
2827 GTTACGCGCGCAACTTGCGGCAATTCCTTCATCAACAGCTGACGTA 2876
830 .....Glulysasnleuasnglnthrlyslleas 839
2877 CGGC.....AATTTGAACGCTCAGGACATTCGCTTATGTCGG 2917
839 .....|||||
839 ngllyalaserthrlyslleusergluarglylphelathrleuylaspl 856
2918 AACTCTTCGCTACCGCGAGCGCAATTAATTAAGCTGGCGGAAATTCGAA 2967
2919 .....|||||
856 ysileasnlylvalarglaserasnleu..... 865
2968 GGCACCTTACACTTGGCTGTCAACATACCGGACGACCGTAGTC 3017
866 .....Serl 867
3018 CGACCAATTCAGCGGTAGTGAAGAAAGAC.....AACACACCG 3058
867 .....|||||
867 aglulneuasnlalthrapslaserglulyslleleasnthrlysl 884
3059 TGTCCGAAATCTTAATTC.....ACCTGCAAAACGAAACACGTGCAT 3102
884 .....|||||
884 euvalserleuaspvalglulysleuvalservalalavalcysasp 900
3103 GCGGCGCATGCGCTTATCAGCTTATCCGCAAGACGCGAGTCCGCT 3152
901 .....|||||
901 Alegly..... 902
3153 GCATTAATCCGCTCAAGAACAGAGCTTCGCAACATCGGC..... 3195
903 .....|||||
903 .....Asnnglycysglulnglulnphleglyasplysglyasnthr 918
3196 .....AAGCGGAGAAACAGAGCC 3216
918 .....|||||
918 ysvalservalglyluleugluala 926
seq_name: swissProt_40:HLXA_SERMA
seq_documentation_block:
ID HLXA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
```

GN SHLA.  
 OS Serratia marcescens.  
 OC Bacteria, Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCB1\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
 RC STRAIN=SN8;  
 RX MEDLINE=88257037; PubMed=3290200;  
 RA Poole K., Schiedel E., Braun V.;  
 RT "Molecular characterization of the hemolysin determinant of Serratia  
 marcescens";  
 RL J. Bacteriol. 170:3177-3188(1988).  
 CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 CC DEFINED.  
 CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
 CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA  
 CC REQUIRES SHLB FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC  
 DR EMBL; M22618; AAA50323.1; -.  
 DR PIR; A28182; A28182.  
 KM Hemolysis; Toxin; Outer membrane; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1608 HEMOLYSIN  
 SQ SEQUENCE 1608 AA; 165078 MW; D6698476FE7DAD51 CRC64;

alignment\_scores:  
 Quality: 213.00 Length: 1609  
 Ratio: 0.282 Gaps: 71  
 Percent Similarity: 46.986 Percent Identity: 18.148

alignment\_block:  
 US-09-303-518D-653 x HLVA\_SERMA ..  
 Align seg 1/1 to: HLVA\_SERMA from: 1 to: 1608

```

172 GCCGAAATTAAGCGAGTTGCGAGCGGCGAAGATATTGAGCTTAA 221
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
23  AAlaAlaSerAlaGlyAlaTyrAla.....AlaGluIleValAlaAlaAsp 37
222 CAAACAAAAGAGGAGTGGTCGCGCAATGCGTACGAAAGCCCGCGATGA 271
    ::  ::|||  |||  ::  ::  |||  |||  |||  |||
37  nGlyAlaAsnGlyProGlyValSerThrAlaAlaThrGlyAlaGlnVal 54
272 TTGATTTTTCGTGGTATCGCGTAAAGCGGTGGCGCATTTGGCGGCGAT 321
    ::|||  ::  ::  |||  |||  |||  |||  |||  |||
54  aAlaSpIleValAlaProAsnGlyAsnGly.....LeuSerHisAsn 67
322 CAATAT.....ATTGAGCGCTGGC 341
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
68  GlnTyrGlnAspPheAsnValAsnGlnProGlyAlaValIleuAsnAsn 84
342 ACATACGGCGGCTATTAACATGTTGATTGGTCGGAGGAGCAACATC 391
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
84  rArgGlnIleGlyLeuSerGlnLeu...AlaGlyIleGlnValAlaAsn 100
392 CCGATCAGCAGCCGCTTTCTTACCAATTTGCAAAAGAAATATATATAA 441
    |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
100  IoAsnIleuGlyGlyArgGlnAlaSerValIleLeu.....AsnIleVal 114
442 GCAGGAGCTAACGGCCATCCTATGCGCGCATATATCATATGCGCGCTT 491

```

```

115  ILeGlyArgAsnProSerLeuHisGlyGlnGlnIle..... 128
492 GCACAAATTTTGCCAGATGCGAAGACCTGTGAGATGACCACTATATAG 541
128 ..... 128
542 ATGGTGAAATATACGCTGATTAATAATATACCGTATGCTGTGCAATC 591
129 .....Phe 129
592 GGAGCAGCAGACAAATATTTGGGCTGTGATGAAGACCAACCAAT..... 636
    |||  ::  ::|||  |||  |||  |||  |||  |||  |||  |||
130  GlyMetAlaAlaAspTyrValIleuAlaAsn.....ProAsnGlyIle 143
637  .AACCGCGAAAGTTTCATATATGCAAGCGCATATTTCTGGCTGTCG 685
143  eSerCysGlnSerCysGlyPheIleAsnThrSerHisSerLeuValIle 160
686 GTGCAATACCTTTGCGCAAAATGATCAGGTGGTCGACAGTCAACTTA 735
    |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
160  aGlyAsnProLeuValGlnAsnGlyValIleuGlnGlyIleSerThrPhe 176
736 GTTACGCAAAAATTAATACATAGCCCATATGTTTTCACCAAGGAGC 785
    |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
177  GlyAsnArgAsnThrLeu...SerLeuAsnGlyThrLeuAsnAlaGlyIle 192
786 CTCATTGGC.....GACAGTGGCTGCACCAATGTTA 817
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
192  yValIleuAspLeuIleAlaProIleAspSerArgGlyGlnValIleVal 209
818 TCTATGATGCCCAAAAGCAAAAGTGGTATTAATGCGGTATTCACAACA 867
    ::  |||  ::|||  ::|||  ::|||  ::|||  ::|||
209  aGlnAsnPheGlyGlnSer.....AsnIle.....LysVal 219
868 GGCACCCCTATATAGCAAAAAGCAATGCGTTCACGATGTTGCTGTAAGA 917
    ::  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
220  ThrSerAlaAlaIleAsnAlaIleSerGlyLeuAsnArgValAlaArgAs 236
918 TTGGTCTATGATGAATCTTTGCGAGATACCAATCAGTATCTTACG 967
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
236  P.....GlyThrValGlnAlaSerGlnGln 245
968 AACCAATCAAAATGGGAATACTTTTTRACGACATATATATAGC... 1014
    |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
245  eProThrAlaLeuAspSerTyrIleuGlySerMetGlnAlaGlyArg 261
1015 .....GCAGGAAA 1022
262  IleAsnIleIleAsnThrAlaGlnGlySerGlyValIleValLeuAlaGly 278
1023 AATGATGCCCAAAATTAACATCTATCTTACTTATATGATTAATAAACAC 1072
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
278  rLeuAsnAlaGlyAspIleuLeuValIleValAlaTyrAspIleArgSerG 295
1073 GAACCGTTCAATGTTAATGTTCTTTATCCGACAGACAGCAAGAACCT 1122
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
295  IuSer...ArgValAspAspAlaSerSerAsnLysAsnGlyGlyAspAsn 310
1123 GTTATCATGCTCAGTGGGCTC.....AACAGTATATGACC 1160
    ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
311  TyrGlnAsnTyrArgGlyGlyIleTyrValAsnAspArgSerSerSerG 327
1161 CAGACTGAT.....AATGACAAATATTTCTTATATGACA 1198
    |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
327  nThrLeuThrArgThrGlyLeuLysGlyLysAsnIleSerLeuValAla 344
1199 AAGGAAGAAGTGATGATCTATACAGCAACATCAACCAAGCGCGGC 1248
    ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
344  sPAsnHisAlaHis.....LeuThrAlaThrSpIleArgGlyGlnAsp 358
1249 GCTTTGATTATTTGAGGTAATTTTACGCTCTGCTAAAAACACGAAC 1298
    |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||

```

```
359 IleThrLeuGlnGlyGlyLysLeuThrLeuAspGlyGlnGlnLeuLysGly 375
1299 GTGGCAAGGCGGCGGCGTTCATATCATAGTATGGCAGTACCGTTACTTGA 1348
      |||||
375 nThrGlnGly.....HisThrAspAspArgTrpPheTyrSerTrpG 389
1349 AAGTAAAGGCGGTGGCAACGACGCGCTGTCAAAATCGGCAAGGCGAC 1398
      : : : : : |||||
389 InTyrAsp...ValThrArgGluArg...GluGlnLeuGlnGlnAlaGly 403
1399 CTGCTGGTTCAGCGCAAGGGAAGAACCAAGCGCTCGCTGCGCTGGCGCA 1448
      ||| |||||
404 SerThrValAlaAlaSerGlySerAlaLysLeuIleSerThrGlnGln 420
1449 CGGTAAAGTCATC...TTAGATCAGCGGCGGACGATCAACGCAAAAAC 1495
      | : : : : |||||
420 PValLysLeuLeuGlyAlaAsnValSerAlaAspArgAlaLeuSerVal 437
1496 AAGCCTTATGAAATCGCTGGTGTACGCGGAGGCGGACGTCGAATG 1545
      : : : : : |||||
437 ysaAlaAlaArgAspValHisLeuAla.....GlyLeuValGluLys 450
1546 AATGCCGATATCATGTTCAACCCGCAAACTATTGGCTTGGCGG 1595
      : : : : : |||||
451 AspLysSerSerGlu.....ArgGly 457
1596 CGGACGCTTGGATTTGACGGGCGATTCGCTTGTCCACCGCATTCAA 1645
      | : : : : |||||
457 yTyrGlnArgAsn...HisThrSerSerLeuArgThrGlyArgTrpSer 473
1646 ATACCGATGAGGGGCGATGATTTGTACCAACATCAAGCAAGAATCC 1695
      || : : : : |||||
473 snSerAspGlnSerGlnSerLeuLysAlaSerGlnLeuArgSerGlnGly 489
1696 ACCGTTACCATTTACAGCAATAAAGATTTTACACACGGG 1737
      : : : : : |||||
490 GluLeuThrLeuLysAlaGlyArgAsnValSerThrGlnGlyAlaLysVal 506
1738 .....AATACACAACTTGGATGACAAAAGAAATGGCTTACA 1777
      : : : : : |||||
506 HisAlaGlnArgAspLeuThrIleAspAlaAspAsnGlnIleGlnValG 523
1778 AC..... 1779
523 LysValGlnLysThrAlaAsnAlaLysAlaValArgAspLysThrSer 539
1780 ...GGTTGGTTGGCGAGAAAGATGCAACCAACGAGGCGGCGCTCAA 1826
      ||| : : : : |||||
540 TrpGlyGlyIleGlyGlyLysAspAsnLysAsnSerAsnAlaArgGly 556
1827 TCTGAATTACCAACCGGAAGAAAGCGGATTCGACTTACTGCTTCCGGCG 1876
      : : : : : |||||
556 ValSerHisAlaSerGlu.....LeuThrSerGlyG 567
1877 GACA...AATTTAAAGGCG.....AATTCACCGCAACCAAGCGCAAA 1917
      ||||| : : : : |||||
567 LysThrLeuArgLysAsnGlnGlnGlnGlnValThrIleThrGly..... 581
1918 CTGTTTTTCAGCGGCGACACGACGCGACGCTTACATCATTTAGGAAG 1967
581 ..... 581
1968 CGGGTGGTCAAAAATGGAAGTATCCCAAGAGAAATCTGTGGGACA 2017
      ||||| : : : : |||||
582 .....SerLysAlaArgGlyGlnLysGlyGlnValThr..... 593
2018 ACGATTGGATCGACGCGACATTTAAAGGGAACCTTCATATTCAGGCG 2067
594 .....AlaThrHisGly 597
2068 GACACAGCGGTG.....GTTTCCGCAATGTGGCAAGTGAAGG 2108
      ||| : : : : |||||
598 GlyLeuArgIleAspAsnAlaLeuSerThrThrValAspLysIleAspAl 614
359 IleThrLeuGlnGlyGlyLysLeuThrLeuAspGlyGlnGlnLeuLysGly 375
1299 GTGGCAAGGCGGCGGCGTTCATATCATAGTATGGCAGTACCGTTACTTGA 1348
      |||||
375 nThrGlnGly.....HisThrAspAspArgTrpPheTyrSerTrpG 389
1349 AAGTAAAGGCGGTGGCAACGACGCGCTGTCAAAATCGGCAAGGCGAC 1398
      : : : : : |||||
389 InTyrAsp...ValThrArgGluArg...GluGlnLeuGlnGlnAlaGly 403
1399 CTGCTGGTTCAGCGCAAGGGAAGAACCAAGCGCTCGCTGCGCTGGCGCA 1448
      ||| |||||
404 SerThrValAlaAlaSerGlySerAlaLysLeuIleSerThrGlnGln 420
1449 CGGTAAAGTCATC...TTAGATCAGCGGCGGACGATCAACGCAAAAAC 1495
      | : : : : |||||
420 PValLysLeuLeuGlyAlaAsnValSerAlaAspArgAlaLeuSerVal 437
1496 AAGCCTTATGAAATCGCTGGTGTACGCGGAGGCGGACGTCGAATG 1545
      : : : : : |||||
437 ysaAlaAlaArgAspValHisLeuAla.....GlyLeuValGluLys 450
1546 AATGCCGATATCATGTTCAACCCGCAAACTATTGGCTTGGCGG 1595
      : : : : : |||||
451 AspLysSerSerGlu.....ArgGly 457
1596 CGGACGCTTGGATTTGACGGGCGATTCGCTTGTCCACCGCATTCAA 1645
      | : : : : |||||
457 yTyrGlnArgAsn...HisThrSerSerLeuArgThrGlyArgTrpSer 473
1646 ATACCGATGAGGGGCGATGATTTGTACCAACATCAAGCAAGAATCC 1695
      || : : : : |||||
473 snSerAspGlnSerGlnSerLeuLysAlaSerGlnLeuArgSerGlnGly 489
1696 ACCGTTACCATTTACAGCAATAAAGATTTTACACACGGG 1737
      : : : : : |||||
490 GluLeuThrLeuLysAlaGlyArgAsnValSerThrGlnGlyAlaLysVal 506
1738 .....AATACACAACTTGGATGACAAAAGAAATGGCTTACA 1777
      : : : : : |||||
506 HisAlaGlnArgAspLeuThrIleAspAlaAspAsnGlnIleGlnValG 523
1778 AC..... 1779
523 LysValGlnLysThrAlaAsnAlaLysAlaValArgAspLysThrSer 539
1780 ...GGTTGGTTGGCGAGAAAGATGCAACCAACGAGGCGGCGCTCAA 1826
      ||| : : : : |||||
540 TrpGlyGlyIleGlyGlyLysAspAsnLysAsnSerAsnAlaArgGly 556
1827 TCTGAATTACCAACCGGAAGAAAGCGGATTCGACTTACTGCTTCCGGCG 1876
      : : : : : |||||
556 ValSerHisAlaSerGlu.....LeuThrSerGlyG 567
1877 GACA...AATTTAAAGGCG.....AATTCACCGCAACCAAGCGCAAA 1917
      ||||| : : : : |||||
567 LysThrLeuArgLysAsnGlnGlnGlnGlnValThrIleThrGly..... 581
1918 CTGTTTTTCAGCGGCGACACGACGCGACGCTTACATCATTTAGGAAG 1967
581 ..... 581
1968 CGGGTGGTCAAAAATGGAAGTATCCCAAGAGAAATCTGTGGGACA 2017
      ||||| : : : : |||||
582 .....SerLysAlaArgGlyGlnLysGlyGlnValThr..... 593
2018 ACGATTGGATCGACGCGACATTTAAAGGGAACCTTCATATTCAGGCG 2067
594 .....AlaThrHisGly 597
2068 GACACAGCGGTG.....GTTTCCGCAATGTGGCAAGTGAAGG 2108
      ||| : : : : |||||
598 GlyLeuArgIleAspAsnAlaLeuSerThrThrValAspLysIleAspAl 614
2109 CGATTGGCATTTAAGCAATACGCCCAACGACTTTTCGTTGCGCA... 2154
      : : : : : |||||
614 a.....ArgThrGlyThrAlaPheAsnIleThrSer 625
2155 .....CCGCATCAAGGCCACACATCTGTACACGT 2184
      : : : : : |||||
625 SerHisLysAlaAspAsnSerTyrGlnSerSerThrAlaSerGluLeu 641
2185 TCGGACTGGGCGGTCTGACAGTTGTACCCGAAAAAACCATTAACGACA 2234
      : : : : : |||||
642 LysSerAspThrAsnLeuThrLeuValSerHisLys.....AspAl 655
2235 TAAAGTATGCTTCA.....TTGAGCAAGACCG 2263
      ||||| : : : : |||||
655 aspValIleGlySerGlnValAlaSerGlyGlyLeuSerValGln 672
2264 ACATCAGAGCGCAATGTCAGCCT...GCCGATCAGCTCATTTAATCTC 2310
      : : : : : |||||
672 erLysThrGlyAsnIleAsnValLysAlaAlaGlnArgGlnGlnAsnIle 688
2311 .....ACAGACTTGGCAGCTCAACGCGCATCTTAGTCAGG 2348
      || : : : : |||||
689 AspGlnGlnLysThrAlaLeu...ThrValAsnGlyTyrAlaLysGlnAl 704
2349 CGGACACACGCACTTACGTT.....ACGCGCACAG 2380
      ||||| : : : : |||||
704 aGlyAspLysGlnTyrArgAlaGlyLeuArgIleGlnHisThrArgAsp 721
2381 CC.....ACCCAAAGCGCACACCTCAGCTTCGTGGCGAAT 2415
      : : : : : |||||
721 erGlnLysThrThrArgThrGlnAsnSerAlaSerSerLeuSerGlyGly 737
2416 GCC.....CAAGCAACATTTAATCAACCCATTAACGCGCAACATC 2459
      : : : : : |||||
738 SerValLysLeuLysAlaGlnLysAspValThrPheSerGlySerLysLe 754
2460 GGGTCGGGACAATGCTTCATTTAATCTAAGCAACACCGCTACAAACG 2509
      : : : : : |||||
754 uValAlaAspLysGlnLysAlaSerValSerGlyAsnLysVal.....S 769
2510 GCAGCTGACGCTTCCGACACGCTAAGCGCAACGTAAGCGAATTCGCCA 2559
      : : : : : |||||
769 erPheLeuAlaAlaAspAspLysThrAlaSerAsnThrGlnGlnThrLys 785
2560 CTCACAGGCAATGTCCTCCTAGCCGATTAAGCAGTATCTTGAATA 2609
      : : : : : |||||
786 IleGlyGlyGly.....PheTyrTyrThrGln 794
2610 C.....AGCGGCTTACCGGAAAAATCAGCGGCGGCAAGATACGCAT 2653
      : : : : : |||||
794 yGlyTyrIleAspLysLeuLysSerGlyValGlnAlaGlyTyrGlnAsnL 811
2654 TACACTTAAAGACAGCGAATGACGCTGCGGCGGCGACGGAATTA... 2700
      : : : : : |||||
811 ysrThrGlnAlaGlnSerSerLysAlaIleThrSerGlySerAspValLys 827
2701 GGCATTTAAACCTTGAC.....AAGCGCAC 2726
      ||||| : : : : |||||
828 GlyAsnLeuThrIleAsnAlaArgAspLysLeuThrGlnGlnGlnAlaG 844
2727 CATTAACATCAATTCGCGCTATGACACGATGAGGAGGCGGCAACCG 2776
      : : : : : |||||
844 nhLysSerValGlyAlaValArgGlnGlnGlnGlnGlnGlnGlnGln 861
2777 GCAAGTGGCGGAGATGCGCGCGCGCGCGCTGCGCGCGCTCCATTAATCC 2826
      : : : : : |||||
861 euAlaAlaAlaAspThrAlaSerThrThrThrLysThrAspValGly 877
2827 GTTACCGCGCACTTCGCGCAGATCCGCTTCAACACGCTGACGTAA 2876
      || : : : : |||||
878 ValAsnIleGlyAlaAsnValAspTyrSerAlaValThrArgProValGly 894
```



seq documentation block:  
 ID PMP6\_CHLPN STANDARD: PRT; 1276 AA.  
 AC 092899; Q9JRM2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable outer membrane protein pmp6 precursor (Polymorphic membrane protein 6).  
 GN PMP6 OR CP00444 OR CP0309.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiae; Chlamydia.  
 OX NCBI\_Taxid=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,  
 RA White O., Hickey E.R., Peterson J., Utterback T., Berry K., Baas S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1 SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)  
 CC (POTENTIAL).  
 CC -1 SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001627; AAD18588.1; -  
 DR EMBL; AE002193; AAF38166.1; -  
 DR EMBL; AP002546; BAA98652.1; -  
 DR PHC1-2DPAGE; Q92899; -  
 DR TIGR; CP0309; -  
 DR InterPro; IPR003368; DUF145.  
 DR InterPro; IPR003357; OMP.  
 DR Pfam; PF02415; DUF145; 2.  
 DR Pfam; PF02385; OMP; 1.  
 KW Outer membrane; Signal; Multigene family; Complete proteome.  
 FT SIGNAL 1 23  
 FT CHAIN 24 1276  
 FT CONFLICT 421 421  
 FT E -> ECGGAILAFIDSGVSKTGSTINNOEVLSTNSA  
 FT AVSGAIVATKCIITGNGSGITPDGNTAGNSGCAITLSTG  
 FT FTILGSGVTFSTNTAKTGALYSKGNLSGNTINLSTG  
 FT NKATGPNSSANOE (IN REF. 1).  
 FT SEQUENCE 1276 AA; 132127 MW; C97FA40662C88E0B CRC64;

alignment\_scores:

Quality: 209.00 Length: 797  
 Ratio: 0.563 Gaps: 36  
 Percent Similarity: 46.530 Percent Identity: 20.828  
 alignment\_block:  
 US-09-303-518D-653 x PMP6\_CHLPN ..  
 Align seg 1/1 to: PMP6\_CHLPN from: 1 to: 1276  
 1570 GACAACTCTATTTCGGCTTCGCGGACGCTTGGATTGAAACGGG.. 1617  
 ||||| ||| ||||| ||||| ||||| |||||  
 75 AsplysSerCysPheThrAsnThrGlyAlaLeuSerPheValGlyAl 91  
 ||||| ||||| ||||| ||||| ||||| |||||  
 1618 ...CATTCGCTTTCCTCCACCGCATCAAAATACCATGAAGGCGCA 1663  
 ||||| ||||| ||||| ||||| ||||| |||||  
 91 AsplysSerCysPheThrAsnThrGlyAlaLeuSerPheValGlyAla 108  
 ||||| ||||| ||||| ||||| ||||| |||||  
 1664 TGATTGTCAACACCAAT..... 1680  
 ||| ||| |||  
 108 lAlaAsnAsnThrAsnThrAlaLeuSerPheSerGlyPheSerLeu 124  
 .....  
 1680 ..... 1680  
 125 LeuIleAspSerAlaProAlaThrGlyThrSerGlyGlyAlaAl 141  
 .....  
 1681 .....CAAGCAAAAGATCCACCGCTTACCATTCACGCAATAAG 1720  
 ||||| ||||| ||||| ||||| ||||| |||||  
 141 ecysValThrAsnThrGlyGlyThrAlaThrPheThrAspAsnAla 158  
 ||||| ||||| ||||| ||||| ||||| |||||  
 1721 ATATTACTACAAACCGCATTAACAACACTGGATAGCAAAAGAAATT 1770  
 ||||| ||||| ||||| ||||| ||||| |||||  
 158 ervalThrLeuGlnIleAsnThrSerGlyAspGlyAlaAlaValSer 174  
 ||||| ||||| ||||| ||||| ||||| |||||  
 1771 GCCTTCAAC.....GGTTGGTTGGCA 1793  
 ||||| ||||| ||||| ||||| ||||| |||||  
 175 AlatySerIleAspLeuAlaGlyThrThrAlaAlaLeuAspG 191  
 .....  
 1794 GAAGATCCAAACCAAGACGGGGGCTCAATCTGAATTACCAACCGG 1843  
 ||||| ||||| ||||| ||||| ||||| |||||  
 191 nasnThrSerThrAsnGlyGly..... 199  
 .....  
 1844 AAGAGCGATCGACTTACTGCTTCGCGGCAAAATTAAACGCG 1893  
 ||||| ||||| ||||| ||||| ||||| |||||  
 200 .....AlaLeuSerThrAlaAsnThrThrValGlnGly 211  
 .....  
 1894 AATATACGCAAAACGCAACTGTTTTCGCGGACCGACACCC 1943  
 ||||| ||||| ||||| ||||| ||||| |||||  
 212 Asn.....SerGlyThrValThrPheSerSerAsnThr 222  
 .....  
 1944 GCAGCGCTACATCATTTAGAACGGGTGGTCAAAATGGAAGGTATCC 1993  
 ||||| ||||| ||||| ||||| ||||| |||||  
 223 ...AlaThrAspIleGlyGlyIleThrSerGlyGly..... 235  
 .....  
 1994 CACAAGGAGAATCGTGTGGCAACGATTGATGACCGCACATTTAA 2043  
 ||||| ||||| ||||| ||||| ||||| |||||  
 236 .....AspSerThrLeuAsp 240  
 .....  
 2044 GCGGAAACTTCATTTACAGCGGCGCAACGGTGTGTTCCCGCAATGT 2093  
 ||||| ||||| ||||| ||||| ||||| |||||  
 241 AlaAsnThr.....GlyValThrPheSerSerAsnThr 252  
 .....  
 2094 TGCCAAAGTGGAAGCGATTGCGCAT.....TTAAGCA 2125  
 ||||| ||||| ||||| ||||| ||||| |||||  
 252 rAlaIleThrGlyAlaThrPheSerSerAspAsnLeuAlaLeuThrG 269  
 .....  
 2126 ATCAAGCCCAAGCAGTTTC.....GGTGGCAGCCG 2157  
 ||||| ||||| ||||| ||||| ||||| |||||  
 269 lAsnThrGlnValLeuPheGlnIleAsnIleThrThrGlySerAlaAla 285  
 .....  
 2158 CATCAAGCCAC.....ACAATCGT..... 2178  
 ||||| ||||| ||||| ||||| ||||| |||||  
 286 GlnAlaAsnAsnProGlnIleGlyGlyAlaIleCysCysIleAla 302

2179 .ACAGTTCGACGTGACGGGTGACA.....A 2206  
 ||| :||| |||||:|  
 302 AThrAlaThrAspLysThrGlyLeuAlaIleSerGlnAsnGlnGlnMetS 319  
 2207 GTTGACGCAAAAAACCATACCGACATGAAAGATGATGCTTCATTGAGC 2236  
 || :|||:| ||| :||  
 319 erPheThrSerAsnThrThrAlaAsnGlyAlaIleTyrAlaThr 335  
 2257 MAGACGACATCAGACGATGACCTGCG..... 2289  
 ||| :|||:| |||||:|  
 336 LysCysThrLeuAspLysAsnThrThrLeuThrPheAspGlnAsnThrAl 352  
 2290 .....GATCAGCGTC 2299  
 332 aThrAlaGlyCysGlyAlaIleTyrThrGlnThrGlnAspPheSerL 369  
 2300 ATTAAATCTCAGAGACTGGCCACACTGCACGCAATCTTAGTCGAGGC 2349  
 :||| :|||:| ||| :||  
 369 euLysGlySerThrGlyThrValThrPheSerThrAsnThrAlaLysThr 385  
 2350 GGAGACAGCCACTATACGGTTACGGCAAGCCACCAAAAGCGCAACT 2399  
 ||| :|||:| |||||:|  
 386 GlyGlyAlaLeuTyrSer.....LysGlyAsnSe 395  
 2400 CAGCTCGTGGCAATGCCCAAGCAACATTT.....AATCAAGCCAGAT 2443  
 |||||:| ||| :|||:| |||||:|  
 395 rSerLeuThrGlyAsnThrAsnLeuLeuPheSerGlyAsnLysAlaThrG 412  
 2444 TAAAGCGCAACATCGCTTGCAACATGCT..... 2475  
 :|||:| |||||:| |||||:|  
 412 LyrProSerAsnSerAlaAsnGlnGlnGlyCysGlyAlaIleLeu 428  
 2476 TCATTAAATCTAAGCAACACCGCTTACAAAAGCGCACTGACGCTTC 2525  
 |||||:| ||| :|||:| |||||:|  
 429 SerPheLeuGlnSerAlaSerValSerThrLysGlyLeuTrpIleG 445  
 2526 CGACAACGCTAAGCAACGTAAGC.....CATTCGCGACTCAAGCGCA 2569  
 :|||:| |||||:| |||||:|  
 445 uAspAsnGlnAsnValSerLeuSerGlyAsnThrAlaThrValSerGly 462  
 2570 ATGTCTCCCTAGCCGATAGCGATATTCAT..... 2601  
 :|||:| ||| :|||:| |||||:|  
 462 LValaIleTyrAlaThrLysCysAlaLeuHisGlyAsnThrThrLeuThr 478  
 2602 TTGGAACAGACCGCGCTTACCGAAAAATCAGGCGCGCAAGATACGAC 2651  
 |||||:| ||| :|||:| |||||:|  
 479 PheAspGlyAsn.....ThrAlaGlnThrAlaGlyGly.....Al 490  
 2652 ATTACACTTAAAGACAGCGAATGAGCGCTGCGCGGCGACGGAATTAG 2701  
 :|||:| |||||:| |||||:|  
 490 aLeuTyrThrGlnThrGlnAspPheThrLeuThrGlySerThr.....G 505  
 2702 GCAATTAAACCTTGACAAGCGCACCATTTACATC..... 2736  
 :|||:| |||||:| |||||:|  
 505 LThrValThrPheSerThrAsnThrAlaLysThrAlaGlyAlaLeuHis 521  
 2737 .....AATTCGCGCTATGACACGATGCGGACGAGCGCAACCGG 2777  
 |||||:| ||| :|||:| |||||:|  
 522 ThrLysGlyAsnThrSerPheThrLysAsnLysAlaLeuValPheSerG 538  
 2778 CAGTGGCGGAGATGCCGCGCGCGCTTCCGCGCTTCATTATTCG 2827  
 :|||:| ||| :|||:| |||||:|  
 538 LysAsnSerAlaThrAla.....ThA 545  
 2828 TTAGCGCGCACTTCGCGCAATCCGTTTACACACGCTGACGGTAAC 2877  
 :|||:| |||||:| |||||:|  
 545 LThrThrThrThrAspGlnGlnGlyCysGlyAlaIleLeuLysAsn 561  
 2878 GGCAATTAAGCGTCAGGACATTCGCTTATGTCGAGACTCTGG 2927  
 561 ..... 561  
 2928 CTACCGCAGCGCAATTGAAGCTGGCGAAGTTCCGAAAGCACTTACA 2977

562 .....:|||||:| :||| :  
 2978 CCTTGGCTGTCAACAATACCGGCAAGAACCCGTAGCTCCGACGCAATG 3027  
 :|||:| ||| :|||:| |||||:|  
 571 erLeuThrLeu.....ThrGlnAsnGlnSerLeuSerPheIleAsnAsn 585  
 3028 ACG.....GTAGTGA 3038  
 ||| :|||:| |||||:|  
 586 ThrAlaLysArgSerGlyGlyIleTyrAlaProLysCysValIleSe 602  
 3039 AGGAAAAGACACACACCGCTGCGAAAATCTTAATTACCTTCGAAA 3088  
 :||| :|||:| |||||:|  
 602 rGly.....SerGlnSerIleAsnPheAspLysAsn 613  
 3089 ACGACACGTCGATGCCGCGCATGGGTTATCAGCTTATCCGCAAGAC 3138  
 :||| :|||:| |||||:|  
 613 hTrAlaGlnThrSerGlyAla.....IleYrSerLysAsn 625  
 3139 GCGGAGTTCGCGCTGCATTAATCCGTCGTAAGAACAGAGCTTCCGACA 3188  
 :|||:| |||||:| |||||:|  
 626 LeuSerIleThrAlaAsnGlyProValSerPheThrAsnAsnSerGly.. 641  
 3189 ACTCGGCAAGCGGGA.....GAAACGAGAGCGCGCT 3220  
 :|||:| |||||:| |||||:|  
 642 ...GlyLysGlyAlaIleTyrIleAlaAspSerGlyLeuSerL 657  
 3221 TGACGGCAAAACAGGCACTTGCAGCCCAACACAGCGGCAAAAGAC 3270  
 :|||:| ||| :|||:| |||||:|  
 657 euGlnAlaIleAspGlyAspIleThrPheSerGlyAsnArgAlaThrGln 673  
 3271 AAGCGCAAAAGCCTTGACGCGCTG...ATTGCGCGCGCGGCAATGCCAC 3317  
 :|||:| ||| :|||:| |||||:|  
 674 GlyThrSerThrProAsnSerIleHisLeuGlyAlaGlyAlaLysIleTh 690  
 3318 CGAAAAGCGCAAAAGCTTGCC.....GAGCGCGGCC 3349  
 :|||:| ||| :|||:| |||||:|  
 690 rLysLeuAlaIleAlaProGlyHisThrIleTyrPheTyrAspProL 707  
 3350 GGCAGCGCAGCGGGGAAATGCGCGCATTTACGAGCGGAGGAAGAA 3399  
 :|||:| ||| :|||:| |||||:|  
 707 hMetGlnAlaProAlaSerGlyGlyThrIleGlnGluLeuValIleAsn 723  
 3400 AAGCGGTGACAGCGGATTAAGACACCGCTTGCGAAGACGCGGAC 3449  
 :|||:| ||| :|||:| |||||:|  
 724 ProValValLysAla..IleValProPro.....GlnP 735  
 3450 GGAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGCGCGCGC 3499  
 :|||:| ||| :|||:| |||||:|  
 735 rLysAsnGlyProIleAlaSerValProValValProValAlaProAla 751  
 3500 GGGATTGCGCAACCGCACCCCAACCGCAACCCCAAC 3538  
 :|||:| ||| :|||:| |||||:|  
 752 .....AsnProAsn 754  
 seq\_name: SwissProt\_40:N100\_YEAST  
 seq\_documentation\_block:  
 ID N100\_YEAST STANDARD: PRT: 959 AA.  
 AC 002629;  
 DT 01-JUL-1993 (rel. 26, Created)  
 DT 01-JUL-1993 (rel. 26, Last sequence update)  
 DT 01-OCT-1996 (rel. 34, Last annotation update)  
 DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).  
 GN NUP100 OR NSP100 OR YKL068W OR YKL336.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 ON NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93054906; PubMed=1385442;  
 RA Wente S.R., Rout M.P., Blobel G.;









[illegible][illegible]

```
2773 .....ACGGCAGTGGCGAGATGCGCGCGCGCGCTGCG 2808
      |||.....:~::~~:
823  leSerValTyrTyrLeuGlyAsnSerThrProThrGluysglYgIAsn 839
2809 CGCGCTTCCCTATTATCCGTTACGCCGCAACTTCGGCAGAAATCCCGTTT 2858
      :~::~~:
840  ThrThrAsnLeu.....ProThrAsnThrThrSer..... 849
2859 CAAACAGCGTGAAGGTAACGCAATTGAACGTCAGGAGACATTCCTCGT 2908
      |||.....:~::~~:
850  .AsnValArgSerAlaAsnAsnAlaLeuAlaGlnAsnAlaProPheAlaG 866
2909 TTATGTCGGAATC.....TTGCGC 2928
      |||
866  lProSerAlaThrProAsnLeuValAlaAlaAsnGlnHisAspPheGly 882
2929 TACCGCAGCGGCAAAATTGAAGTCGGCGGAAGTCCGGAAGGCACTTAC 2978
      :~::~~:
883  ThrIleGluSerValPheGluLeuAlaAsnArgSerLysAspIleAspTh 899
2979 CTGGCTGTCAACATATACCGGCAACGACCCGTAAGTCTCGAGCAATTGA 3028
      |||.....:~::~~:
899  rLeuTyrAlaAsnSerGly..... 905
3029 CGGTAGTGAAGGAAAGACAAACACACCGCTGCCGAAATCTTAATTTC 3078
      :~::~~:
906  .....AlaGlnGlyArgAsp.....LeuLeuGln 913
3079 ACCCTGCAAAAGCAACACGTCATCGCGCGCATCGCTTATACCTTAT 3128
      |||.....:~::~~:
914  ThrLeuLeuIleAspSerHisAspAlaGlyTyrAlaArg...GlnMetI 929
3129 CCGCAAGACGCGGAGTCCGCTGCATTAATCCGGTCAAGAAACAAGAGC 3178
      |
929  e.....AspAsnThrSerThrGlyGluI 937
3179 TTTCGCAACAACCTCGGCAAGCGGAGAAACAAGACCGCTTGAACGGCA 3228
      :~::~~:
937  lThrLysGlnLeuAsnAlaAlaThrThrThrLeuAsnAsnIleAlaSer 953
3229 AAACAGGCAACACTTCCGCCCAAAACAAGCGCGGAAAGACACACGGCA 3278
      :~::~~:
954  LeuGlnHisLysThrSerSerLeuGlnThrLeuSerLeuSerAsnAlaMe 970
3279 AAGCTTGAACGG...CTGATTCGCGCGCGGCAATGCCACCGCAAAAG 3325
      :~::~~:
970  tIleLeuAsnSerArgLeuValAlaAsnSerArgAlaGlnHisAsnAla 987
3326 CAGAAAGTGTTCGCAACCGCGCGCGGAGGAGCGGCGGAAATGCCGCG 3375
      :~::~~:
987  lAspSerPheAlaGln..... 992
3376 ATTATGCAAGCGGAGAGAGAGAAAAACGGGTGCGAGCGGATAAAGAC 3425
      :~::~~:
993  .....ArgLeuGlnAlaLeuLysAspG 1000
3426 CGCCTTGGCAACAGCCGGAAGCGGAAACCGCGCGCTACACCGCT 3475
      :~::~~:
1000  nLysPheAlaSerLeuGlnSerAla..... 1008
3476 TCCCCCGCGCGCGCGCGCGGATTTGCCGCAACGCAAGCCCA 3525
      :~::~~:
1008  ..... 1008
3526 CCGCAACCCCAACCGCAGCGGACCTGATCAAGCGTTATGCCAATAGCG 3575
      :~::~~:
1009  .....AlaGluValLeuTyrGlnPheAlaProLys.. 1018
3576 TTGTAGTGAATTTCCGCGCAGCTCAACAGCGTTTCCGCTACAGAGCG 3625
      :~::~~:
1018  ..... 1018
3626 AATTGGACCGGCTTTTCCGCAAGACCGCGCAACGCCGTTTGGACAAGC 3675
```

```
1019 .....TyrGluLysProThrAsnValThrPalaAsn 1028
      :~::~~:
3676 GGCATCCGGGACACCAACACTACCTTCGCAACATTTCCGCGCTACCG 3725
      :~::~~:
1029 AlaIleGlyGlyThrSerLeuAsnAsnGlyGlyAsnAlaSerLeuTyrG 1045
3726 CCACAAACCGACCTGGCGCAATCGGTATGCAAAAAACCTC..... 3768
      :~::~~:
1045  yThrSerAla.....GlyValAspAlaTyrLeuAsnGlyG 1057
3769 .....GGCAGCGGCGCGCGCTTGTAGCGCGCATTCCTTTTCG 3798
      :~::~~:
1057  lValGluAlaIleValGlyGlyPheGlySerTyrGly...TyrSerSer 1072
3799 CACAACCGGACCGGAAACACCTTCGAGACGCGCATTCGCAATCGGCA 3848
      :~::~~:
1073  PheAsnAsnGlnAlaAsnSerLeuAsnSerGlyAlaAsnAsnThrAsn 1089
3849 GCTTCCCGCAGGTCGCTTTCGGGCAATACGCAATCGCAGGTTGCA 3898
      :~::~~:
1089  eGlyValTyrSerArgIlePheAlaAsnGlnHisGluPheAspPheGlu 1106
3899 TCGGCATCAGCGCGCGGCTTTGTAGCGCGCATTCGAGAGCG 3948
      :~::~~:
1106  lGlnGlyAlaLeuGlySerAspGlnSerSerLeuAsnPheLysSerAla 1122
3949 ATCAGAGCAAAATCCGCGCGCGCTGCATTAACGCAATTCAGGCA 3998
      :~::~~:
1123  LeuLeuArgAspLeuAsnGlnSerTyrAsnTyrLeuAlaTyrSerAla 1139
3999 ATACCGGCAAGCTTCGCGC.....GGATTCG 4024
      :~::~~:
1139  aThrArgAlaSerTyrGlyTyrAspPheAlaPhePheArgAsnAlaLeu 1156
4025 GCATCGAACCGCACATCGCGCAGCGCTATTGTCGCAAAAGCGAT 4074
      :~::~~:
1156  aLleuLysProSerValGlyAlaSerTyrAsnHisLysGlySerThrAsn 1172
4075 TACGATACGAAACGTCATATGCGCACCGCGCGCTTCATTCACCG 4124
      :~::~~:
1173  PheLysSerAsnSerAsnGlnValAla..... 1181
4125 CTACCGCGCGGCGCATTAAGCAGATTTATTCATTCAAACCGCGCAC 4174
      :~::~~:
1182  .LeuLysAsnGlySerSerSerGlnHisLeuPheAsnAlaSerAlaAsn 1198
4175 TTTCATCAGCGCTTATTGAGCCTGTCTATACCGCATTCGCGCTCCG 4224
      :~::~~:
1198  aGluAlaArgTyrTyr.....TyrGlyAspThrSer..... 1208
4225 AAAGTCCGAACCGCGTCATATACCGCGGTATTCGCGCAG.....GATT 4268
      :~::~~:
1209  .....TyrPheTyrMetAsnAlaGlyValLeuGlnGlnPheAlaAsn 1223
4269 CGGCAAAACCGGCACTGCGGAA.....TGGGCGGTAAACCGCGGAA 4309
      :~::~~:
1223  eGlySerSerAsnAlaValSerLeuAsnThrPheLysValAlaAsnAla 1240
4310 TCAAAGGTTTCAGCGCTGCTCCCTCCACGCTGCCCGCAAGGCGGCA 4359
      :~::~~:
1240  lAsn.....ProLeuSerThrHisAlaArgValMetMetGlyGlyGlu 1254
4360 TTGAGACCGCAGCAGCGCGCGCATCAATTAATGAGCTAC 4398
      :~::~~:
1255  LeuLysLeuAlaLysGluValPheLeuAsnLeuGlyPhe 1267
```

seq\_name: Swissprot\_40:120K\_RICRI

seq\_documentation\_block:

ID 120K\_RICRI STANDARD; PRT; 1300 AA.  
AC P14914;  
DT 01-APR-1990 (Rel. 14, Created)

DT 01-FEB-1996 (Rel. 33, last sequence update)  
 DT 01-MAR-2002 (Rel. 41, last annotation update)  
 DE 120 kDa surface-exposed protein.  
 GN P120.  
 OS Rickettsia rickettsii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=783;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R.  
 RX MEDLINE=90136087; PubMed=2515418;  
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;  
 RT "Cloning, expression and sequence analysis of the gene encoding the  
 RT 120 kD surface-exposed protein of Rickettsia rickettsii";  
 RL Mol. Microbiol. 3:1579-1586(1989).  
 CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS  
 CC -1- RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A  
 CC -1- S-LAYER WITH HEXAGONAL SYMMETRY.  
 CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES  
 CC -1- CONFERRING ANTIGENICITY TO THE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 DR EMBL: X16353; CAA34402.1; -  
 DR PIR: S07575; S07575.  
 DR InterPro: IPR003858; rompa\_rompb.  
 DR Pfam: PF02708; rompa\_rompb; 1.  
 KW Antigen; Glycoprotein; Cell wall; S-layer.  
 FT CARBOHYD 7  
 FT CARBOHYD 66  
 FT CARBOHYD 86  
 FT CARBOHYD 103  
 FT CARBOHYD 147  
 FT CARBOHYD 268  
 FT CARBOHYD 330  
 FT CARBOHYD 375  
 FT CARBOHYD 415  
 FT CARBOHYD 424  
 FT CARBOHYD 430  
 FT CARBOHYD 436  
 FT CARBOHYD 444  
 FT CARBOHYD 515  
 FT CARBOHYD 547  
 FT CARBOHYD 593  
 FT CARBOHYD 655  
 FT CARBOHYD 698  
 FT CARBOHYD 710  
 FT CARBOHYD 799  
 FT CARBOHYD 800  
 FT CARBOHYD 826  
 FT CARBOHYD 844  
 FT CARBOHYD 861  
 FT CARBOHYD 879  
 FT CARBOHYD 920  
 FT CARBOHYD 926  
 FT CARBOHYD 1116  
 FT CARBOHYD 1128  
 FT CARBOHYD 1140  
 FT CARBOHYD 1146  
 FT CARBOHYD 1211  
 SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

alignment\_scores:

Quality: 203.50 Length: 1269  
 Ratio: 0.371 Gaps: 53  
 Percent Similarity: 43.262 Percent Identity: 19.228  
 alignment\_block:  
 us-09-303-518D-653 x 120K\_RICRI ..  
 Align seg 1/1 to: 120K\_RICRI from: 1 to: 1300  
 1225 AGCAACATCAACCAAGCGCGCTTGTATTTGAGGATATTTTAC 1274  
 72 SerAspProGlyAsnThrAlaGlyValIleThrPheAspAlaAsnGlyTh 88  
 1275 G.....GTCTCGCTTAAACCAAGAA 1297  
 88 rleuGluSerAlaSerAlaAspAlaAsnValAlaValThrAsnAsnIleT 105  
 1298 CG.....TGGCAAGCGCGGCGCTT..... 1317  
 105 hrAlaIleGluAlaSerGlyAlaGlyValAlaGluIleuSerGlyThrHis 121  
 1318 .....CATATCATGA 1328  
 122 AlaAlaGluLeuArgLeuGluAlaGlySerIlePheGlyLeuAlaAs 138  
 1329 TGGCAATACGCTTACTTGGAAAGTAACGCGTGGCAACGACCGCGTGT 1378  
 138 pGlyThrValIleAsnGlyLysValAsnGluThrAla..... 150  
 1379 CCAAAATCGCAAAAGCAGCGCTCTGTTCAAGCCAAAGGCAAAACCA 1428  
 151 ..LeuValGly..... 153  
 1429 GGCTCGTCAAGCGTGGC.....GACGTGAAGTCATCTTGA 1466  
 154 GlyAlaLeuAlaAlaGlyThrIleThrLeuAspGlySerAlaThrIleTh 170  
 1467 TCAGCAGCGCGGACATCAAGGCAAAACCAAGCCTTACTGAATTCGGCT 1516  
 170 rGlyAspIleGlyAsnAlaGlyLysAlaAlaIleuGluThrL 187  
 1517 TGTGTACG...GGCAGGGGAGCGTGCACTGAATCGCATATCATGTC 1563  
 187 euAlaAsnAspAlaLysLysThrLeuThrLeuGlyLysAlaAsn..... 201  
 1564 AACCCGCAAACTATTCGCTTCGGCGGCGACGTTGGATTTG.. 1611  
 202 .....IleIleGlyAlaGlyLysIleThrIleAspLeuG 213  
 1612 ....AACGGCATTCGCTTCGTTCCACCGCATTCAAAT..... 1647  
 213 naLaAsnGlyThrIleLysLeuThrSerThrGluAsnAsnIleValIv 230  
 1648 .....ACCGATGAAGCGCATGATTCATC 1674  
 230 alAspPheAspLeuAlaIleAlaThrAspIleThrGlyAlaValAspAla 246  
 1675 CACATCAAGCAAGAACATCCACCGTACCATTCAGGCAATTAAGATAT 1724  
 247 SerSerLeuThrAsnAlaGluThrIleThrIleAsnGlyLys..... 261  
 1725 TACTACACCGCGCAATRACACCAACTTGGATACCAAAAAAGAAATTCCT 1774  
 261 ecGlyThrIleGlyAlaAsnAsn..... 268  
 1775 ACAAGGTTGTTGGGAGAGAAAGATGCACCAACCAAGGCGCGCTC 1824  
 269 .....LysThrLeuGlyGluPhe 274  
 1825 AATCTGAATTAACCAACCGAAGAGCGATGCATTTACTGCTTCGCG 1874  
 275 AsnIleGly.....SerSerLysThrValLeuSerAsnG 286

```
1875 CGGACAAATTAAAC.....GGCAAT..... 1896
      |||
286 yasnValAlaIleasnGluLeuValIleGlyAsnAspGlyAlaValaGlnP 303
1897 .....ATCGGCAACA.....AAC 1911
303 heAlaHisAspThrTyrLeuIleThrArgThrThrAsnAlaIleGlyGln 319
1912 GGCAACGTGTTTTC.....AGCGAGACGACACACCGCAGC 1949
      |||
320 GlyLeuIleIlePheAsnProValValaAsnGlyThrThrLeuAlaI 336
1950 CTACAACTATTGGAAGCGGTGTCAAAATGGAAGGTATCCGCAAG 1999
      |||
336 acGlyThrAsnLeuGlySerAlaThrAsnProLeuAlaGluIleAsnPhg 353
2000 GAGAAATCGTGTGGACACGATGTGATGACCGCAGCATTTAAAGCGGA 2049
      |||
353 LyserylSglYValaAsnValaAspThrValLeuAsnValGlyGlyVal 369
2050 AACTTCATATT..... 2061
370 AsnLeuTyrAlaThrAsnIleThrThrThrAspAlaAsnValGlySerPh 386
2062 .....CAGGCGGACAAAGCGGTGTTCCGCAATGTTGCCAAG 2101
386 eValPheAsnAlaGlyGlyThrAsnIleValSerGlyThrValGlyGly 403
2102 TGGAAGCGATTTGGCATTTAAGCAATCAGCCCAAGCATTTGCGTGC 2151
      |||
403 InclIneLy..... 405
2152 GCACCGCATCAAGCCACAAATCTGTACAGTTCGACGAGCGGTCT 2201
      |||
406 .....AsnLysPheAsnThrValAlaLeuGluAsnGlyThrThr.... 418
2202 GACAACTTACCAGAAAAACCATTAACGACGATTAAGTATGCTTCAT 2251
419 .....ValLysPheL 422
2252 TGAGCAAGACGACATCAGAGCAATGTACAGCTTCCGATACCGCTCAT 2301
      |||
422 euGlyAsnAlaThrPheAsnGlyAsnThrThrIleAlaIleAsnSerThr 438
2302 TTAATCTCACAGACTTCGCACTCAACGCAATCTTAAGTGGCGG 2351
439 LeuGlnIleGlyGlyAsnTyrThrAlaAspGlyValAlaSerAlaAspG 455
2352 AGACACGCACTATACGTTACGCGACAGCCACCAAAAGCGCAACTCA 2401
455 Y.....ThrGlyIleValaGluPheValaAsnThrGlyProIleT 468
2402 GCCTGTGGCAATGCCCAACA.....ACATTAATCAACCC 2439
      |||
468 hrValThrLeuAsnLysGlnAlaIleProValaAsnAlaLeuLysIle 484
2440 ACATTAAAGCAACACATCGGCTTGGACATGCTTCATTTAATCTAAG 2489
      |||
485 ThrValSerGlyProGlyAsnValaValIleAsnGluIleGlyAsnAlaG 501
2490 CAACAC.....GCCGTACAAAAGCGCACTC 2515
      |||
501 yAsnHisHisGlyAlaValaThrAspThrIleAlaIlePheGluAsnSerS 518
2516 TG..... 2517
518 euGlyAlaValaValaPheLeuProArgGlyIleProPheAsnAspAlaG 534
2518 ...ACGCTTTCGACACGCTAAGGCAAGTAAAGCACTTCGCACTCAA 2564
535 AsnThrMetProLeuThrIleLysSerThrValGlyAsnLysThrAlaL 551
2565 CGGC.....AATGTCCTCCCTAG 2581

      |||
551 sGlyPheAspValProSerValValaValLeuGlyValaLysSerValIleA 568
2582 CCGATAAGGCAATTCATTTTGAAAACAGCCGCTTACCGGAAAAATC 2631
568 laAspGlyGlnValIleGlyLysAspGlnAsn.....Ile 579
2632 AGCGGC.....GGCAAGATACGGCATTTACACTTAAAGACACGA 2672
580 ValGlyLeuGlyLeuGlySerAspAsnGlyIleIleValaAsnAlaThr.. 595
2673 ATGACGCTGCCGTGGCGACGCAATTAAGCAATTTAAACCTTGACAACG 2722
596 ...ThrLeuTyrAlaGly.....IleSerThrLeuAsnAsnGlnG 609
2723 CCACCATTAACCTCAATTCGCCCTATGCACACGATGCGGCA..... 2763
609 lYThrValThrLeuSerGlyGlyValaProAsnThrProGlyThrValTyr 625
2764 GGCGCGCAAAACCGGACAGTGGCGAGATGCGCGCGCGCGCTTGGCGCG 2813
626 GlyLeuGlyThrGlyIleGlyAlaSer..... 634
2814 TTCCTATTATCGCTTACGCGCCCACTTCGCGCAATTCGCTTCAACA 2863
635 LysPheLysGlnValThrPheThrThrAspTyrAsnAsnLeuGlyAsnI 651
2864 CGCTGACGCTAAC..... 2877
651 leIleAlaThrAsnAlaThrIleAsnAspGlyValaThrValThrThrGly 667
2878 .....GGCAATTGAA 2888
668 GlyIleAlaGlyIleGlyPheAspGlyLysIleThrLeuGlySerVala 684
2889 CGGTACGAGCAATTCGCTTATGTGGAATCTTCGCTACGCGCGC 2938
684 nGlyAsnGlyAsnValaArgPheAlaAspGlyIleLeuSerAsnSerThrs 701
2939 GCAATTGACGCTGCGGAAAGTTCGAGAGCACTTACCTTGCGT... 2985
701 erMetIleGlyThrThrLysAlaAsnAsnGlyThrValaThrThrLeuGly 717
2986 .....GTCACAACTACCGGCAACGA.....CCGTAAGTCTCGA 3020
718 AsnAlaPheValaGlyAsnIleGlyAspSerAspThrProValaIleSerVa 734
3021 GCAATTGACGCTAGTGAAGGAAAGCAACACACCGCTTCGGAATC 3070
734 lArgPheThr.....GlySerAspSer.....GlyAlaGlyL 745
3071 TTAATTTCACCTCGCAAAAGCAACACGTCATGCCGCGCATGGCTTAT 3120
745 euGlnGlyAsnIleTyrSerGlnValIleAspPheGlyThrTyrAsnLeu 761
3121 CAGCTTACCGCAAAAGAC.....GGCAGTTCGCGCTGCA 3155
762 GlyIleValaLysSerAsnIleIleLeuGlyGlyGlyThrThrAlaIleAs 778
3156 TAAATCGGTCAAAAGCAAGACTTTCGACAAACTCGGCAAGCGGAG 3205
778 n.....GlyLysIleAspL 783
3206 AAACAGAGGCGCGCTTGAGGCAAAACAGGACACACTGCGCGCAACAA 3255
783 euValThrAsnThrLeuThr..... 789
3256 CAGCGGAAAAAGAACACGCGCAAGCTTGACGCTGATGCGGCGG 3305
790 .....PheAlaSerG 793
3306 GCGCAATGCCAGAAAAGGACAGAAAGTGTGCGGAAACGCGCGGCGAG 3355
```

```

793 yThrSerThrTrpGlyAsnAsnThrSerIleGluThrThrLeuThrLeuA 810
3356 CAGCGGGGGAAGATCCCGCATTTATGAGCGGAGAGAGAAAAACGG 3405
810 laAsnGlyAsnIleGlyHisIleValIleLeuGlu..... 821
3406 GTGACGGGGGATAAAGACACCGCCTTGCGAAACAGCGCGAAGCGAAGC 3455
822 .....GlyAlaGlnVal 825
3456 CCGGCGGCTACACCGCCTTCCCGCGCGCGCGCGCGCGGATT 3505
825 laSnThrThrThrGly..... 831
3506 TGCCCGCAACCGACGCCCAACCGCAACCGCACGCGACTGATC 3555
831 ..... 831
3556 AGCCGTATGCAATAGCGGTTTGTAGTATTTCCGCGCATCAACG 3605
832 .....ThrTh 833
3606 CGTTTCCGCTACAGAGCAATTGACCGCGTGTTCGCGAAGACGCC 3655
833 rThrIleYsValGlnAspAsnAlaAsnPheserGlyThrGlnT 850
3656 GGAACGGCGTTTGACAAAGCGCATCGGGACACCAACACTACGCTG 3705
850 hTyrThrLeuIleGlnGlyAlaIleArgPheAsnGlyThrLeuGlySer 866
3706 CAAGATTTCGCGCTACCGCGCAACCAACGCACTGCGCAATCGTAT 3755
867 ProAsnPhE..AlaValThrGlySerAsnArgPheValAsnYrSerLeuI 883
3756 GCAGAAAAAATCTCGGACGGCGCGTGGCATCTCTGTTTTCGCAAC 3805
883 laArg.....AlaIleAsnGlnAspTyrValIleThrArgThrAsn 896
3806 GGACGGGAACACCTTCGACGCGGCACTCGGACGCGCTGCG 3855
897 AsnAlaGlnAsnValValThr.....AsnAspIleAl 907
3856 CACGTCGCGTTTTCGCGCAAT.....ACGGCAT 3884
907 aAsnSerProPheGlyAlaProGlyValAspGlnAsnValThrThp 924
3885 CGGACGTTTCGATCGCATCGCGCGCGCGGCTTTAGTACG... 3931
924 heValAsnAlaThrAsnThrAlaAlaTyrAsnAsnLeuLeuAlaYs 940
3932 .....GCAGCTTTCAGACGGCATCATGAGCGCAAAATCGCG 3966
941 AsnSerAlaAsnSerAlaAsnPhEValGlyAlaIleValThrAspThse 957
3967 CGCGCGCTCGCATTCAGCATTCAGCGCAATACCGCGAGTTTCGG 4016
957 rIleAla.....IleThrAsnValGlnLeuAspLeuAlaYsApIleG 972
4017 CGGATTCGCGATCGAACCGCATCGCGCAACGCGTATTTCTGCCAA 4066
972 laIleGlnLeuGlyAsnArgLeuGlyAlaLeuArgTyrLeuGly..... 986
4067 AAGCGATTACGGATACGAAACGTCATATGCGACCCCGCGCTGCA 4116
987 .....ThrProGluThrAlaGluMetAlaGlyProGluAlaGlyAl 1000
4117 TTCACCGCTACCGCGCGCGCATTAAGCGAGATTATTCATAACCGCG 4166
1000 aIleSerAlaIleValAlaIleGlyAspGluAlaIleAspAsn..ValAl 1016
4167 GCAACACATTCATCAGCGCTTATTTGAGCTGTCTATACCATGCC. 4215
1016 aTyrGlyIleTrpAlaYsProPhe.....TyrThrAspAlaH 1029

```

```

4216 .....GTTCCGCAAGTCCGACGCG... 4239
1029 tsGlnSerLysGlyGlyLeuAlaGlyTyrLysAlaYsThrThrGly 1045
4239 ..... 4239
1046 ValValIleGlyLeuAspThrLeuAlaAsnAspAsnLeuMetIleGlyAl 1062
4240 .....GTCAATACCGCGCTATTTGCGCGAGATTTCGCAAAA 4276
1062 aAlaIleGlyIleThrLysThrAspIleLysHisGlnAspTyrLysLys. 1078
4277 CCGCATGCGCGAATGCGCGTAAACGCGCAATCAAGATTTCACGCTG 4326
1079 .....GlyAspLysThrAspAlaYsAlaYsPheSerPhe 1089
4327 TCCTCCACGCTGCGCGCGCAAGGCGCGCATTTGGAACGCGACAG 4376
1090 SerLeuTyrGlyAlaGlnGlnLeuValLysAsnPhEAlaGlnGlyse 1106
4377 CGCG 4380
1106 rAla 1107

seq_name: SwissProt_40:PMPB_CHLMU
seq_documentation_block:
ID PMPB_CHLMU STANDARD; PRT; 1672 AA.
AC Q9PUY2;
DT 16-OCT-2001 (Rel. 40, Created)
DR 16-OCT-2001 (Rel. 40, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpB precursor (Polymorphic membrane
DE protein B).
GN PMPB OR TC0694.
OS Chlamydia muridarum.
OX Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_Taxid=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / N199;
RX MEDLINE=20150253; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClary G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE002336; AAF39510.1; -.
CC TIGR: TC0694; -.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14
FT CHAIN 15 1672 POTENTIAL.
FT CHAIN 15 1672 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SO SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8BD6 CRC64;

```



## alignment\_scores:

Quality: 203.50 Length: 1139  
 Ratio: 0.398 Gaps: 55  
 Percent Similarity: 44.864 Percent Identity: 20.720

## alignment\_block:

US-09-303-518D-653 x PMPB\_CHLMU ..

Align seg 1/1 to: PMPB\_CHLMU from: 1 to: 1672

```

706 AATGATCAGAGTGTGGCAGACGACCTTAGAGCGCAAAATAATTAACA 755
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 AsnGlyAsnGlyAla.....AspSerGlnThrProSerH 257
756 TAACCATATGCTTTTACACAGAGAGCGCTCATTTGGGACAGGCGCT 805
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 slvspro.....GlySerGlyAlaIleTyrAlaThrGlyAla 270
806 CACCAATGTTTATCTATGATCCCAAAAGCAAAAGTGAATTAATGAG 855
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
270 sp...LeuThrIleSerAspSerGlnGluIleValPheSerValAsnLys 285
856 GTATTGCAACAGCAACCCCTATAGCAAAAAGCAAT...GGCTTCCA 902
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
286 AlaSerLysAspGlyGlyAlaIlePheAlaGluLysAsnValSerPheG 302
903 G.....CTAGTCGTAAGATGGTCTCATGATGA... 933
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
302 uAsnIleThrThrLeuLysValGlnAsnGlnAlaGluGluLysGlyG 319
934 ....ATCTTGGCTGAGATPACCATTCAGTATTCAGAACCCATCAA 978
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
319 LysIleTyrAlaSerGlyAspLeuSerIle.....Gln 330
979 AATGGAATATAC...TTTAAACGACATATAAT... 1013
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 SerSerLysGlnSerLeuPheAsnSerAsnThrSerLysGlnGlyGly 347
1014 CGCA.....GGAATAATCGATGCAAAACAT...AAACACTAT 1048
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
347 yAlaLeuTyrIleGluGlyAsnValAspPheLysAspLeuGluGluIle 364
1049 CTCCTACTATAGATTAAACACGACGCTTCAATGTTTAACTTCT 1098
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 rGileLysTyrAsn...LysSerGlyThrPheGluThrLysLysValThr 379
1099 TTATCC..... 1104
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
380 LeuSerLeuProGluAlaGlnThrAsnLysSerSerValThrAlaLys 396
1105 .....GAGACAGCAAGAGAACCTGTTATTCATG 1132
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
396 rGlnSerGlyProAsnThrThrProThrProThrProValThrAlaL 413
1133 CTGACAGTGGGTCAACAGTTATCGACCCAGCTGAATAATGAGAAAT 1182
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
413 ysgGlyGlyLys.....LeuTyrThrGlnLysAsn 422
1183 ATTTCCTTATTGACAAAGAAAGTGAATGATGATCTTACAGCAAC.. 1230
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
423 LeuSer...IleSerAsnIleThrGlyIleIleGluIleThrAsnAsnLys 438
1231 .ATCAACCAAGCGCGCGCGCTTATATTGAGGCTAATTTTACGGTCT 1279
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
438 sAlaThrAspValGlyGlyGlyAlaTyrValLysGlyThrLeuThrCysL 455
1280 CG.....CCTAAACCAAGCAAGAGCGTGGCAAGC 1308
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
455 yAspSerHisArgLeuGlnPheGlnLysAsnSerSerGlnLysLysGly 471
1309 GCGGCGCTTCATATCATGATGAGCAGTACCTTACGTAAGTAAGCG 1358
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
472 GlyGlyLeuTyrThrGluAsp.....ThrIleThr..... 481

```

```

1359 CGTGGCAAAACGACCCCTGTCCAAATCGCAAAAGCAGCGCTGTGTTTC 1408
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
482 .....LeuSerAsnLeuThrGlyLysThrLeuPheGln 493
1409 AA.....GCCAAGGGGAAACCAAGCTCGGTACGCTGGCGACGCT 1452
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
493 LuAsnThrAlaLysGluGluGlyGlyLysLeuTyrIleGlnLysAsp 509
1453 AAGTCATC.....TTAGATCAG..... 1470
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
510 LysThrLeuThrMetThrGlyLeuAspSerPheCysLeuIleAspAsnTh 526
1471 .CAGCGGACGATCAAGCAAAACCAAGCCTTAGTGAATTCGGC... 1515
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
526 rSerAlaThrHisGlyGlyAlaTyrValThrLysGlnIleSerGlnT 543
1516 .....TTGGTCAGCGGCGAGGGAGCGTGCACACTGAATGCC 1551
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
543 hTyrThrSerAspValGluGluPheProGlyIleThrProValHisGly 559
1552 GATAATCAGTTCACCCGACAAACTATTCGCGCTTTCGCGCGCA.. 1599
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
560 GluThrIleIleSerGlyAsnLysAlaThrGlyLysSerGlyGlyAla 576
1600 .....CGTTGGATTGAACGGGCGCATTCGCTTCT 1630
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
576 lCysThrLysHisLeuValLeuSerAsnLeuGlnThrIleSerIleSer 593
1631 TCACCCGATTCAAATAATCCGATGAGGGCGGATGATGTCACACCAAT 1680
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
593 LuAsnPheAlaSerGlnLysnclGlyGlyAlaCysThrCysProAspAsn 609
1681 CAAGCAAGAATCCACCCTTACCATTAAGCAAT..... 1716
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
610 PheProAlaProThrAlaSerThrProSerThrAsnGlnThrAlaLap 626
1717 .....AAGAT.....ATTACTACAAACCGCA 1738
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
626 OlyAspAspLysAspPheLeuIleAspTyrValValSerThrThrIle 643
1739 ATACCAACACTTGATAGCAAAAGAAATGCGCTCAACGCTGTGTT 1788
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
643 sPlyAsnLysAlaThrLysLysGlyAla.....GlyValTyr 655
1789 GCGCAGAAAGATGCAACCAACGACGGCGCTCATCTGAAT..... 1833
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
656 AlaLysLysAlaLysLeuSerArgIleAspGluLeuAsnIleSerAsp 672
1834 .....TACCAACCGGAGAG 1849
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
672 nAlaLagLingluThrGlyGlyPheCysCysThrGlnSerLeuGlu 689
1850 CGGATCCGACTTACTCTTTCGCGGCAACAAATTTAAAGGC..... 1893
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
689 euAspThrIleAlaSerLeuSerValThrLysAsnLeuAlaGlyGlu 705
1894 .....AATATACGCAAAACCAAGCA 1916
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
706 GlyGlyLysLeuHisAlaLysThrLeuAsnIleSerAsnLeuLysSer 722
1917 ACTGTTTTTCAGC..... 1929
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
722 yLeuSerPheSerAsnAsnThrAlaAsnSerSerSerThrGlyValAla 739
1929 ..... 1929
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
739 hThrAlaThrThrSerGlnSerProThrValSerSerPheLeuProArg 755
1930 .....GGCAGACGACACCGGACGCTA 1952
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
756 AlaThrAlaLysSerSerProAlaProAlaGlnThrThrProThrTyrAl 772

```

```

1953 CAATCATTTAGGAGC.....GGTGTGCAAAA 1981
    :: :::::
772 aGlyValValGlyGlyAlaIleTyrGlyGlnThrValSerPheSerLys 789
    :::::
1982 TGGAGGATCCACAGAGCAAAATCGTGTGGACACAGATTGGATGCAC 2031
    :::::
789 ySerGlyLeuGlyGln.....PheThrGluAsnSerAlaIleAsp 802
    :::::
2032 GCGCATTTAAAGCGGAAACTTCATATTCGAGCGGCAAGACGGGTGT 2081
    :::::
803 AsnThrProSerSerProSerLeuAsnValGlnGlyGly...AlaIleTyr 818
    :::::
2082 TTCGCGCATTTGTCGCAAGAGTGGAGGCGATGGCATTTAAAGCAATTCAC 2130
    :::::
818 rAlaLysThrSerLeuSerIleGluAlaGluAsProSerThrSerTyr 835
    :::::
2131 .....GCCCA 2136
    :::::
835 aLpHeSerLysAsnSerValSerThrGlyLysAlaGlnThrThrGlyGln 851
    :::::
2137 GCGATTTTGGTGTGCGACCGCATCAAGCCACACATCTGTACAGCTTC 2186
    :::::
852 lIleAlaGlyGlyAlaIleTyrSerProSerValThrLeuAsnGlyGlnThr 868
    :::::
2187 GGACTGGAGCGGT.....CTGACAAGTTGTACCGAAA 2218
    :::::
868 rValPheSerGlyAsnSerAlaSerMetAlaThrThrAsnProProSerG 885
    :::::
2219 AATCATTTCCGACATAAAGATGATGCTTCATTCATGCAAGCCGACATC 2268
    :::::
885 lYThrSerProLysAspThrIleGlyGlyAlaIleAlaGlyThrThrIle 901
    :::::
2269 AGAGCGAATGTCAGCTTGCGATCAGCGTCATTTAAATCTCAGCAGAT 2318
    :::::
902 .....SerLeuSerLysThrSerThrPheSerGluAsnThrAlaAspLe 916
    :::::
2319 TCCGACACTCAACGCGCATCTTAGTCAGCGGACAGACAGCAC..... 2361
    :::::
916 uGlyAlaAlaIleGlyThrLeuSerGlyLysSerSerSerAsnLeuThrG 933
    :::::
2362 .....TATACGTTACGGCGCAAC...GCC 2382
    :::::
933 lAluLysIleThrLeuSerAsnGlySerPheThrPheGluLysAsnLysAla 949
    :::::
2383 ACCCAAAACGCG.....AACCTCAGCTCGTGGCGCAATGC 2417
    :::::
950 AsnLysArgGlyValIleTyrAlaProSerValSerIleLysGlyAsn... 965
    :::::
2418 CCAAGCAATTTAATCAAGCCACATTAAGCGCAACACATCGGCTTGGG 2467
    :::::
966 .AsnIleThrPheAsnGln.....AsnThrSerThrHisa 977
    :::::
2468 ACAATGCTCATTTAATCTAAGCAACACGCGCTA.....CAAAACGCG 2511
    :::::
977 spGlySerAlaIleTyrPheThrLysAspAlaThrIleGluSerLeuGly 993
    :::::
2512 AGT...CTGACGCTTTCGACACAGCTTAAGCGCAAGCAATGCGCTCCG 2558
    :::::
994 SerValLeuPheThrGlyAsnAsnValThrAlaGlnGlnAlaSerSerTh 1010
    :::::
2559 ACTCAACGCG.....AATGTCCTCAGCGCATAGAGAGTAT 2596
    :::::
1010 rAlaThrGlyGlyGlnThrThrAsnThrThrAsnTyrGlyAlaAlaIleP 1027
    :::::
1027 heGlyAspProGlyThrGlnThrThr.....Asp 1036
    :::::
2647 AGCGCATTTACTTAAGACAGCAAGCAATGAGAGCTGCGCTGGGACGGA 2696
    :::::
1037 ThrThrLeuLysLeuIleAlaSerSer..... 1045
    :::::
2697 ATTAGCAATTTAAACCTTGACAC.....GCCACATTTACACTCA 2737
    :::::
1046 .....GlyAsnIleThrPheSerAsnAsnSerGlnAsnThrAlaThrAsnA 1061
    :::::
2738 ATTCCGCGCTATGCACAC..... 2754
    :::::
1061 snProLalaThrLysPheCysSerIleSerGlyTyrValLysLeuThrLeu 1077
    :::::
2755 GATCGCGGAGCGCGGCAACCGGCACTGCGGAGATGCGCGCGCGCGCG 2804
    :::::
1078 GluAlaIleGlnGlyLysThrIleSerPhePheAspSerIleArgThrse 1094
    :::::
2805 TTCGCGCGGTTCCTATTTATTCGCTTACGCGCGCAACTTGGGCAAAATCC 2854
    :::::
1094 rThrLysLysThr.....GlyGlnAlaGlnAsn 1104
    :::::
2855 GTTTCACACGCGTACGCGTAACGCGCAATTAAGACGGTGC..... 2895
    :::::
1104 eTyrGluThrLeuAspIleAsnLysThrGluAsnSerAsnThrTyrAla 1120
    :::::
2896 GAAACATTTCCGCTTATGTGCGAATCTTGGCTTACGCGCGCAAAAT 2945
    :::::
1121 GluThrValLeuPheSerSerGluLeuIleGlnValLysSerTyrValPr 1137
    :::::
2946 GAACTGGCGGAAAGTTCCGAGGCACTTACACTTGGCTGTGTAACAA 2995
    :::::
1137 oGlnAsnValValLeuHisAsnGlyThrLeuValLeuLys...LysAsnA 1153
    :::::
2996 CCGGCAACGACCGCGTACGCTGAGCAATTCAGCATGAGGAAAGAAA 3045
    :::::
1153 lAgIleuHisValValSerPheGluGln.....Lys 1163
    :::::
3046 GACACACACCGCTGTGCGAAATCTTAATTTACCTCGCAAAACAGACA 3095
    :::::
1164 GluIleThrLeuSerLeuIleMetGluProGlyAlaValLeuSerAsnGlnAs 1180
    :::::
3096 CGTGATGCGCGCGCATGGCGTTATGAGTTATCGGCAAAAGAGGAGAG 3145
    :::::
1180 nIleAlaAsnGlyAlaLeuAlaIleAsn..... 1189
    :::::
3146 TCCGCTGCAATTCGCGTCAAGAACACAGAGCTTTCGCAAACTCGGC 3195
    :::::
1189 ..... 1189
    :::::
3196 AAGCGGGAGAACACAGAGCGCGCTTGCAGCGCAAAACAGCACAACTTGC 3245
    :::::
1190 .....GlyLeuThrIleAspLeuSerSerLeuGln 1199
    :::::
3246 CGCCAAACACAGCGGGAAGAAAGACACAGCGCAAGCTTGACGCGCTGA 3295
    :::::
1199 yAlaProGlnThrGlyGlyValPheSerProProGlnLeu..... 1212
    :::::
3296 TTGCGCGCGGCGCAATGCACCGCAAAAGGAGAAAGTGTTCGCAACG 3345
    :::::
1213 .....ArgIleValAlaThrThr 1218
    :::::
3346 GCGCGGAGAGAGCGGGAAGAAATGCGGCATATGACGCGGAGAGAGA 3395
    :::::
1219 SerAsnSerGlyGlyGlyGlyValGly..... 1228
    :::::
3396 GAAAAAAGCGGTGCGAGGAGATAAGACACCGCTTGGCGAAACAGCGCG 3445
    :::::
1229 ....GlyTyrValThrAlaSerLysAsnLeuSerAlaIleAspProThr 1244
    :::::
3446 AAGCGGAAACCGCGCG 3462
    :::::
1244 aAlaAlaThrAsnPro 1249
    :::::
seq_name: SwissProt_40:PM20_CHLPLN
seq_documentation_block:
ID PM20_CHLPLN STANDARD; PRT; 1723 AA.
AC Q9Z812; Q9KZC1; Q9RB59;
DT 16-OCT-2001 (Rel. 40, Created)

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable outer membrane protein pmp20 precursor (Polymorphic membrane protein 20).  
 GN PMP20 OR CEN0540 OR CP0212.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83558;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CM1029.  
 RA MEDLINE=99206606; PubMed=10193388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RA MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RA MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shirai T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RA MEDLINE=20298986; PubMed=10839753;  
 RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,  
 RA Takeuchi H., Nishida J., Shibata K., Fujinaga R., Yoneda H.,  
 RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,  
 RA Ishii K., Shiba T., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of outer membrane protein genes omp and pmp in the whole genome sequences of Chlamydia pneumoniae isolates from Japan and the United States.";  
 RL J. Infect. Dis. 181 Suppl 3:S524-S527(2000).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES) (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_email_to_license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001638; A018680.1; -  
 CC EMBL: AE002181; A038082.1; ALT\_INT.  
 CC EMBL: AP002547; BAA98746.1; -  
 CC EMBL: AB033817; BAA85968.1; -  
 CC PHC1-2DPAGE; O92812; -  
 CC TIGR: CP0212; -  
 CC InterPro: IPR003368; DUF145.  
 CC InterPro: IPR003357; OMP.  
 CC Pfam: PF02415; DUF145; 1.  
 CC Pfam: PF02385; OMP; 1.  
 KW Outer membrane; Signal; Multigene family; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 1723 PROBABLE OUTER MEMBRANE PROTEIN PMP20.  
 FT DOMAIN 94 97 POLY-SER.  
 FT DOMAIN 140 144 POLY-SER.  
 FT DOMAIN 147 160 POLY-SER.  
 FT CONFLICT 1134 1134 I -> V (IN REF. 1).  
 SQ SEQUENCE 1723 AA; 179594 MW; 4507D5B80E3EF01 CRC64;

## alignment\_scores:

Quality: 203.50 Length: 985  
 Ratio: 0.461 Gaps: 49  
 Percent Similarity: 44.772 Percent Identity: 21.218

## alignment\_block:

US-09-303-518d-653 x PM20.CHLPN ..

Align seg 1/1 to: PM20.CHLPN from: 1 to: 1723

679 CTGCTGCGTGGCAATACCTTGCACAAATGGATGAGTGGTGCACACT 728  
 ||| .....  
 434 LeupheserGlyAsnThr.....AlaAsnYsserGlyGlyGlyIleTy 448  
 .....  
 729 CACTTATAGGTAGCGAAATAATTAACATAGCCCATATGCTTTTATACCA 778  
 .....  
 448 rAlaValGlyGlnValThrLeuGluAspIleAlaAsnLeuYsserMetThr 465  
 .....  
 779 CAGGAGGCTCATTTGGCGACAGTGGCTGCACCAATGTTATCTATGATGCC 828  
 .....  
 465 snAsnThrCysIysGlyGlyGlyGlyAla.....IleTy...Thr 477  
 .....  
 829 CAAAGCAAAAGTGGTATTAATGAGG...GATTTGCAACAGCAACCC 875  
 .....  
 478 IysIysAlaLeuThrIleAsnMetIysAlaIleLeuThrThrPheSerG 494  
 .....  
 876 CTATATAGCAAAAGCAATGCTTCACACTAGTTCGTAAAGATTGTTCT 925  
 .....  
 494 yAsnThrSerThrAspAsnGly..... 501  
 .....  
 926 ATGATGAATCTTCTGCT.....GGAGATACCATTCAGTATTTCTACGAA 969  
 .....  
 502 ..GlyAlaIlePheAlaValGlyIleThrLeuSerAspLeuValGlu 517  
 .....  
 970 CCACATCAAAATGGGAATACTTTTACGACGCAATATATATGCGCAGG 1019  
 .....  
 518 ValArg.....PheSerIysAsnIysThrGly.... 526  
 .....  
 1020 AAAAATCGATGCCAAACATAACACTATTCCTACCTTAT...AGATTAA 1066  
 .....  
 527 .....AsnTySerAlaProIleThrIysAla 536  
 .....  
 1067 AACACGACGCTTCATGTTATGTTATGTTTATTCGACGACGACGACAA 1116  
 .....  
 536 IAsnAsnThrAlaProVal.....ValSerSerSerThrThrAlaAla 550  
 .....  
 1117 GAACCTGTTTATCATGCTGCA..... 1137  
 .....  
 551 SerProAlaValProAlaAlaAlaAlaAlaProValThrAsnAlaIle 567  
 .....  
 1138 .....GTGGGGCTACACA 1150  
 .....  
 567 sGlyGlyAlaLeuTySerThrGluGlyLeuThrValSerGlyIleThr 584  
 .....  
 1151 GTTATCGACCCAGACTGATGATAGGAAATAATATTCCTTATTCACAAA 1200  
 .....  
 584 er.....IleLeuSerPheGluAsnGlnGlyGlnAsnGln 596  
 .....  
 1201 GGAAGAGCT.....GA 1211  
 .....  
 597 GlyGlyGlyAlaTyValThrIleThrPheGluCysSerAspSerHisar 613  
 .....  
 1212 ATGATACTTACCGCAAC.....ATCACCAAGGCGCGGGGTTTGT 1255  
 .....  
 613 gLeuGlnPheThrSerAsnIysAlaAlaAspGlyGly...GlyGlyLeu 629

```

1256 ATTTGAGGTAATTTTACGGCTCG..... 1281
1259 YTCYSGLYASPAVALThrLeuThrAsnLeuThrGlyLysThrLeuPhe 645
1282 CCTAAAAACAAGAACGTGGCAAGCGCGGCGCTCATATCAGTATG 1331
646 GlnGluAsnSerSerGlnLysHisGlyGlyLeuSerLeuAlaSerGI 662
1332 CAGTCCCTTCTTGGAAGTAAGACGGCGTGGCAACACCGCCGTCGA 1381
662 YLysSerLeuThr.....MetIrrs 669
1382 AAATGGCAAAAGCACGCTGCTGTTCAAGCCAAAGGGAAACCAAGC 1431
669 erLeuGluSerPheCysLeuAsnAlaAsnThrAlaLysGluAsnGly 685
1432 TCGGTACGCGTGGCGGCGGTAAAGTCATCTTA..... 1464
686 GlyAlaAsnValProGluAsnIleValLeuThrPheThrProTh 702
1465 ...GATACGACGGCGGACGATCA.....GGCAAAAACAAG 1498
702 rProAsnGluProAlaProValGlnGlnProValTyrGlyGluAlaLeu 719
1499 CCTTAGTGAATCGCTTGGTACGCGGAGGCGGCGTCAATGAT 1548
719 alThrGlyAsnThrAlaThrLysSerGlyGlyIleTyrThrLysAsn 735
1549 GCCGATATCAGTCAACCCGACAAACTATTTCCGCTTGGCGGCGG 1598
736 Ala.....AlaPheSer..... 739
1599 AGCTTTGATTTGAACGGCATTCGCTTCCACCGCATTCAAATA 1648
740 .....AsnLeuSerSerValThrPheAsp.....GlnAsnT 750
1649 CCGATGAA.....GGGCGATGATTTGCACCAACATCAAGACAA 1689
750 hSerSerGluAsnGlyAlaLeuLeuThrGlnLysAlaAlaAspLys 766
1690 GAA.....TCCACCGTACCATTCAGGCAATAA 1718
767 ThrAspCysSerPheThrTyrIleThrAsnValAsnIleThrAsn.. 782
1719 AGATATTACTACACCGCATTAACAACAATCTG..... 1752
783 .....ThrAlaThrGlyAsnGlyGlyIleAlaGlyLysAlaH 797
1753 .....GATAGCAAAAATA 1764
797 lAspHeaParGlyLeaSPAsnLeuThrValGlnSerAsnGlnAlaLysLys 813
1765 .....GAAATTCGCTACAA 1778
814 GlyGlyGlyValTyrLeuGluAspAlaLeuIleLeuGlyLysValIleTh 830
1779 CGGTGGTGGCGAAGAAAGATCAACAACAAAGAAAGCGCGGCTC... 1824
830 rGlySerValSerGlnAsnThrAlaThrGluSerGlyGlyIleTyrA 847
1825 .....AATCTGAATTACCAACCGGAAGACGGATCGCACTTACTG... 1866
847 lAlysAspIleGlnLeuGlnAlaLeuProGlySerPheThrIleThrAsp 863
1867 .....CTTCCGGCGGACAAATTTAAAGCGCAATAT 1898
864 AsnLysValGluThrSerLeuThrThrSerThrAsnLeuTyrGly.... 878
1899 CACGCAAAACAAACGCAAACTGTTTTCAGCGGCGACACCGACACGCGACG 1948
879 .....GlyGlyIleTyrSerSerGlyAlaValThrLeu.... 889

```

---

```

1949 CCTACAAATCATTTAGAACGGGTGTCAAAAATGAGATATCCACAA 1998
890 .....ThrAsnIleSerGlyThrPheGlyIleThrGly 900
1999 GGAGAAATGCTGTGGACAACGATTCATCGACCCGACATTTAAAGCGGA 2048
901 AsnSerValIle.....AsnThrAlaThrSerGI 910
2049 AAATTCATATTTCAAGGCGGA.....CAAGCGGTGTTCGCGCAATG 2092
910 nasPAlaAspIleGlnGlyGlyIleTyrAlaThrThrSerLeuSerI 927
2093 TTGCCAAATGGAAGCGATGTCATTTAAGCAATPACCCCAAGAGAT 2142
927 leaSnGlnCysAsnThrProIleLeuPheSerAsnSerAlaAla... 942
2143 TTCGGTGTCCACCGCATCAACACACAAATCTGACGTTGCGACTG 2192
942 ..... 942
2193 GACGGGTGTGACAAGTTGACGAAAAAACATTACCGACGATAAAGTG. 2241
943 .....ThrLysThrSerThrThrLysGlnIleA 953
2242 .....ATTGCTTCTTGACGAAGACGACATTCAGAGGCAT... 2277
953 laGlyAlaIlePheSerAlaIleValThr...IleGluAsnAsnSer 968
2278 .....GTCAAGCTTGGCGATCAGCTCATTTAATCTCAGAGACT 2318
969 GlnProIleIlePheLeuAsnAsnSerAlaLysSerGluAlaThrAl 985
2319 TGCCACACTCAACGGCAATCTTGTGACGCGGA.....GACACGC 2359
985 alAlaThrAlaGlyAsnLysAspSerCysGlyAlaIleAlaAlaAsn 1002
2360 ACTATACGTTTACCGGCAAC.....GCCACCAAAAGCGCACTCAGC 2403
1002 erValThrLeuThrAsnAsnProGluIleThrPheLysGlyAsnTyrAla 1018
2404 CTCGTGGCAATGCCAAGACATTTAATCAAGCCACATTAACGGCAA 2453
1019 GluThrGlyGlyAlaIleGlyCysIleAsp.....LeuThrAsnGly 1033
2454 C.....ACATCGGCTTCGACATGCTTCACTTAATCTAAGCA 2491
1033 rProProArgLysValSerIleAlaAspAsnGlySerValLeuPheGlnA 1050
2492 ACAAGCGCGTACAAAC..... 2508
1050 spAsnSerAlaLeuAsnArgGlyAlaIleTyrGlyGluThrIleAsp 1066
2509 .....GGCAGCTGACGCTTCCGACACGCTAAGCAACGAT 2546
1067 lIleSerArgThrGlyAlaThrPheIleGlyAsnSerSerLysAspGI 1083
2547 AAGCATTCGCCACTCAACGCAATGCTCCCTAGCCGATTAAGCAATAT 2596
1083 ySerAlaIleCysSerThrAlaLeuThrLeuAlaProAsnSerGlnL 1100
2597 TCCATTTGAAGAACCGGCTTACCGGAAATCAGCGGCGGCAAGAT 2646
1100 euIlePheGluAsnAsnLysValThrGluThrAlaThrThrLysAla 1116
2647 ACGCATTTACACTTA.....AAAGACAGCGA 2672
1117 SerIleAsnAsnLeuGlyAlaAlaIleTyrGlyAsnAsnGlnThrSerAs 1133
2673 ATGACGCTGCCGTCGCGACGGAATTAAGCAATTAACCTTGACAACG 2722
1133 pIleThrIleSerLeuSerAlaGluAsnGlySerIlePhePheLysAsnA 1150
2723 CCACCATTAACACTCAATTCGCTATGACACAGATGCGGCGGCGCAA 2772

```





693 laLeuHisLeuLysAspGlySerGlyAlaGluLeuProIleAsnProGly 709  
1294 GAACGCTGGCAAGCGCGGC..... 1314  
710 ProThrTyrLysAsnIleGlyAspGlyAlaLysArgAsnTyrSerPhe 726  
1315 .....GTTCAATATCAGTGAATGGCAGTACCGTTACTTGAAAGTAA 1354  
726 eLysIleIleAlaTyrAsnLysAspGlyAlaIle.....AlaS 740  
1355 AGCGCGTGGCAACGACCGCTGTCCAAATATGGCAA..... 1392  
740 eProlAlaIleThrProAlaLeuProAspIleAlaArgProLysAsnVal 756  
1393 ...GGACCGCTGTGTTCAAGCCAAAGGGGAAACCAAGGC..... 1431  
757 ThrGlyTyrLeuTyrThrAsnThrLysSerSerGlnThrGlyTyrValAs 773  
1432 .....TCGCTCAGCGTGG 1444  
773 nLeuIleTrpGluLysValGlnAsnAlaLysGlyTyrLysValAsnIleTr 790  
1445 GCGACGGTAAAGTCATCTTAGATCAGCAGCGGACGATCAAGCAAAAA 1494  
790 yTrAsnGlyLys.....GluTyrGlnSerPheAspValGlyAspAla 803  
1495 CAAGCCTTTAGT.....GAAT 1511  
804 AspHisTrpThrThrGlnAsnLysAsnIleTrpThrSerGlnIle 820  
1512 CGGCTTGTCAGCGGAGGAGCGGTGCACATGTAATCCGTAATATCAG 1560  
820 e.....LysAlaGlySerTyrLysLeuHisThrAspGlyLysG 833  
1561 .....TTCAACCCCGACCAACTGTAT..... 1581  
833 LysGlyLeuLeuAlaLeuAspProSerProValTyrAsnAsnAlaAsnGly 849  
1582 .....TTGCGCTTTCGCGGAGCGGACGTTTGATTT 1610  
850 AsnTyrLysGlyLysLysAsnTyrSerPheThrLeuValAlaTyrAspAl 866  
1611 CAAGCGGCAATTCGCTT.....TCGTTCCACGCA 1639  
866 aAsnGlyGluThrIleProThrAlaProPheAsnProThrPheHis 881  
1640 TTCAAATAATACGATGAGGGGCGATGATTCACACCAATACAGACAAA 1689  
882 .....GluGlyAlaGluPheLeuGly...ThrGluIleTyr 892  
1690 GAATCCACCGTTACCAT.....ACAGGCAA 1715  
893 TrpSerIleIleAspIleProSerGlyLysLeuAsnGlyAlaThrGlyLys 909  
1716 T.....AAGATATTACTACAAACGCGCAATAAC..... 1743  
909 nValIleValAsnGluAspLeuSerIleAspGlyArgGlyProGlyL 926  
1744 .....AACACTGTGATGCAAAAAAAGAAATGCC 1773  
926 euGlyLeuSerArgThrTyrAsnSerLeuAspSerSerAspHisLeuPhe 942  
1774 TACAACGGTGGTTGGC..... 1791  
943 GlyGlnGlyTrpTyrAlaAspAlaGluThrSerValIleSerThrAspG 959  
1792 .....GAGAAAGATGCACAC.....A 1807  
959 yGlyAlaMetTyrIleAspGluAspAlaThrThrHisArgPheThrLysL 976  
1808 AAACGAAACGGCGGCTCAATCTGAATATCAACAGG..... 1842  
976 ySalAspGly.....ThrTyrGlnProProThrGlyValTyr 988

1843 .....GAAGAAGCGGATCGCATTTACTGCTTCCGCGGAC 1880  
989 LeuGluLeuThrGluThrAlaAspGln...PheIleLeuLysThrLysAs 1004  
1881 AAATTTAAACGGCAATATACGCAAAACAGCAACGCTGTTTTCAGCG 1930  
1004 pGlnThrAsnAlaTyrPheAsnLysLysGlyGlyLysLeu.GlnLysVal 1020  
1931 GCAGACCGACACCGCACCGCTACATC..... 1957  
1021 ValAspGlyHisAsnAsnAlaThrValTyrThrTyrAsnAspLysAsnG 1037  
1958 .....ATTAGAGAGGGGTGTCAAAAATGGAAGTATCCACA 1997  
1037 nLeuThrAlaIleThrAspAlaSerGlyArgLysLeuThrPheThrTyrA 1054  
1998 AGGAGAAATCGTG..... 2010  
1054 spGluAsnGlyHisValThrSerIleThrGlyProLysAsnLysVal 1070  
2011 .....TGGACAACGATTGGATC.....GACCG 2033  
1071 ThrTyrSerTyrGluAsnAspLeuLeuLysValThrAspThrAspG 1087  
2034 CACATTTAAAGCGGAAACTTCATATTCAGGGCGGCAAGCGGTGTT 2083  
1087 yThrValThrSerTyrAspTyrAspSerGluGly..... 1098  
2084 CCCGCAATGTTGCCAAAGTGAAGGCGATGGCATTTAGCAATCAGCGC 2133  
1099 ..ArgLeuValLysGlnTyrSerAlaAsn.....SerThrGluAla 1111  
2134 CAACGACTTTTCGGTGTGCGACCGCATCAAGCAGCAACATCTGTACAGC 2183  
1112 LysProValPheThrGlnTyrGlnTyrSerGlyHisArgLeu..... 1125  
2184 TTCGACTGACGCGGTGTGACAAAGTTGTACGAAAAACCATAC... 2229  
1126 .....GluLysAlaIleAsnAlaL 1132  
2230 .....GACGATAAAGTATGCTGCTCA 2250  
1132 yLysGluThrTyrValTyrSerTyrAspAlaAspLysLysThrLeu 1148  
2251 TTGACGACAGACCGACATCAGA.....GG 2273  
1149 MetThrGlnProAsnGlyArgLysValGlnTyrGlyTyrAsnGluAlaG 1165  
2274 CAAT...GTCAGCCTGCGATCAGCTCAT...TTAAATCTCAGAGAC 2317  
1165 yAsnProIleGlnValIleAspAspAlaGluGlyLeuLysIleThrTha 1182  
2318 TTGCCACACTCAACGCGCAATCTTAGTCAGCGGAGACACGCACTATACG 2367  
1182 snThrLysTyrGluGlyAsn..... 1188  
2368 GTTACGGGCAACGCCACCCCAAAAGCGCAACCTCAGCTCTGGGCAATGC 2417  
1189 .....AsnValValGluAspValAspProAsnAspAlaGlyThrG 1202  
2418 CCAAGCAACATTTAAT...CAAGCCACATTAAGCGCAACACATCGGCT 2464  
1202 yLysAlaThrGlnSerTyrGlnTyrAspLysAspGlyAsnValThrSerV 1219  
2465 CGGACAATGCTTCATTTAATCTAAGCAACACGCGCTACAAAAGGCACT 2514  
1219 aLysAspAlaTyrGlyThrGluThrTyrGlnTyrAsnLysAsnAsnAsp 1235  
2515 CTGACGCTTCGAGCAACGCTAAGCAACGTAAGCATTCGCGCACTCA 2564  
1236 ValThrLysMetLysAspThrGlnGlyAsnValThrAspIleAlaTyrAs 1252

```

2565 CGGC.....AATGTCCTAGCGATAC.....2589
      ||| ||| ||| |||
1252 pclyleuaspalaValSerGluThrAspGlnSerGlySerSerSera 1269
2590 ..GCAGTATTCATTGAAAAACCGCTTACCGGAAAAATACAGCGC 2637
      ||| ||| ||| |||
1269 laalavalThr.....AspLysTYGlyAsnGlnIleGlnSer 1281
2638 GGCAGGATACGGCATAC.....TTAAAGAC.....2667
      ||| ||| ||| |||
1282 SerLysAspLeuSerAlaSerThrAsnIleLeuLysAspGlySerPheG 1298
2668 .....AGCAATGAGCGCTCCGCGGCGCAGCAATTAAGCAATT 2707
      ||| ||| ||| |||
1298 ualaglInlySerGlyTrpAsnLeuThrAla.....1308
2708 TAAACCTGACAGCGCAGCATTTACATCAATTCGGCTATCGACAGCAT 2757
      ||| ||| ||| |||
1309 ..SerLysAspArgArgLysIleSerValIleAla.....Asp 1320
2758 GCGGAGGCGCGCAACCGCAGTGCAGATGCGCGCGCGCGCGCTTC 2807
      ||| ||| ||| |||
1321 LysSerGlyValLeuSerGlySerLysAla.....1330
2808 GCGCGCTCCCTATTATCCGTTACCGCGCACTTCGGCAGAACCCGTT 2857
      ||| ||| ||| |||
1346 yfSerSerAlaThrGlnThrValGlnLeuGluProAsnThrThrThr 1362
2908 TTTATGTCGAGACTCTCGCGTACCGCAGCGCAATTTGAG.....CT 2951
      ||| ||| ||| |||
1363 Leu.....SerGlyLysIleLysThrAspLe 1371
2952 GCGGAAATTCGAGAGCATTACACCTTGGCTGTCACAAATACCGGCA 3001
      ||| ||| ||| |||
1371 ualalys...SerArgAlaThrPheAsnIleAspLeuArgAspLysAspG 1387
3002 AGCAACCCCTAGTCTC.....GAGCAATTGACGTAAGTGAAGAA 3045
      ||| ||| ||| |||
1387 InLysArgIleGlnThrPheAsnGlnLysSerAlaLeuAlaGlyLys 1403
3046 GACACACACCGCTGTCGAAATCTTAATTCAC.....3081
      ||| ||| ||| |||
1404 ...AsnAspTrpThrLysArgGlnIleThrPheThrProAlaAsnAl 1419
3082 .....CTGCAAAACGACACGTCGAT.....3102
      ||| ||| ||| |||
1419 agLysAlaValAlaValLysMetGluValAspHisLysAspLysGlyL 1436
3103 ..GCGGCGCATGGCTTATCAGCTTATCCGCAAGAGCGGAGTTC... 3147
      ||| ||| ||| |||
1436 ysgLysAlaThrPheAspGluValGlnLeuGlnLysGlyLysValSer 1452
3148 CCCTTCATATCCGTCGCAAGACAGCTTTCGCAAAATC.....3192
      ||| ||| ||| |||
1453 SerSerTyrAsnProValGlnAsnSerSerPheThrSerAlaThrGluAs 1469
3193 .....GCGAAGCGCGGAGAAACAGAGCCCGCTTACGCCCA 3229
      ||| ||| ||| |||
1469 nTrpAsnValSerGlyAlaSerValAspSerGluGlnLysPheAsnAsp 1486
3230 AACAGGCACACTGCGCGCAACAAACAGCGGAAAAAGACAAAGCGGCA 3279
      ||| ||| ||| |||
1486 spValSerLeuLysAlaAlaThrSerAlaSerGlnAlaLysSerVal 1502
3280 ACCCTTGACGCGCTGATTGGCGCGCAATTCACGCAAGAAAGCGAGA 3329
      ||| ||| ||| |||
1503 ThrLysGlnThrValVal...LeuGlyGlnSerAlaAsnAspLysProVa 1518
3330 AAGTGTGCGGAACCGCGCGCAGGCGGGAATAATGCGGCGCATTA 3379

```

```

1518 lTyrLeuThr.....LeuThrGlyMetSerLysAlaSerSerV 1531
      ||| ||| ||| |||
3380 TGCAGCGCGGAGAGAAAGAAAAA...CGGGTCGAGCGGAGTAAAGACAC 3426
1531 alLysPheThrAspGluLysAspTyrSerLeuGlnAlaAsnValThrTyr 1547
3427 GCCTTGCGAAGACGCGCGAAGCGGAAACCGCGCGGTACACCGCTT 3476
      ||| ||| ||| |||
1548 Ala.....AspGlySerThrGlyIleTyrAsnAlaLysPhe 1559
3477 CCCCCCGCGC.....CGCGCGCGCGCGCGGATTTGCGGCAAC 3514
      ||| ||| ||| |||
1559 eProSerGlyThrGlnGluTrpAsnArgAlaAlaValIleProLysT 1576
3515 CGAGCGCCCAACCGCAACCCCAACCGCAGCGGACCTGATACGCGTAT 3564
      ||| ||| ||| |||
1576 hLysPro.....1578
3565 GCCAATAGCGGTTTGAGTGAATTTCCGCCACGCTCACAGCGTTTCC 3614
      ||| ||| ||| |||
1579 .....IleAsnLysVal.....1582
3615 CGTACAGAGCAATTTGACCGCGCTTTGCCGAAGACCGCGCAACGCG 3664
      ||| ||| ||| |||
1583 .....AspLysIleLeuPheGlnLysSerAlaThrGlyThrV 1596
3665 TTGGACACACGCGCATCCG.....3684
      ||| ||| ||| |||
1596 alTrpPheAspPheArgLeuIleGluGlySerLeuLeuThrLysSer 1612
3685 .....GACACCAACACACTAC.....CGTTCGCAAGATTTCCGCG 3719
      ||| ||| ||| |||
1613 ThrTyrAspSerAsnGlyAsnTyrValThrLysGluGluAspGluLeuG 1629
3720 CTACCGCAACAAACGACCTGCGCCAAATCGGTATGCAAGAAATAC... 3765
      ||| ||| ||| |||
1629 yTyrAlaThrSerThrAspTyrAspGluThrGlyLysLysThrSerGluT 1646
3765 .....3765
1646 hrAspAlaLysGlyGluLysThrThrTyrThrTyrAspGlnAlaAspGln 1662
3766 .....CTGCGACGCGGCGCGCTGCGCATCTGTTTCGCA 3800
1663 LeuThrAsnMetThrLeuSerAsnGly...ThrSerIleLeuHisSerTy 1678
3801 CAACCGGACCGGAAAC 3816
1678 rAspLysGluGlyAsn 1683
seq_name: swissprot_40:VAC3_HELPY
seq_documentation_block:
ID VAC3_HELPY STRAND: PRT: 1310 AA.
AC Q48253;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxId=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TX30A;
RX MEDLINE=9535366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummur M.K., Blaser M.J.,
RA Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vacA types with cytotoxin production and

```



[illegible]

```
1783 ....TGTTGGGAGAAAGATGCACCAAAAGCAAGCGGCTAATC 1828
      |||
      |||
      |||
516 rAlaHisPheLysAsnIleAspIaThrLysSerAspAsnGlyLeuAsnIle 533
      |||
      |||
      |||
1829 TGAATTACCAACCGAAGAGCGGATTCGCTTACTGCTTCCGGCGGGA 1878
      |||
      |||
      |||
533 hrser.....ThreusPheSerGlyVal 541
      |||
      |||
      |||
1879 ACAAT.....TTAAACGGCAATATCAGCAAAACAA 1910
      |||
      |||
      |||
542 ThrAspLysValAsnIleAsnLysLeuThrThrAlaIaThrAsnValAs 558
      |||
      |||
      |||
1911 CGGCAAACTGTTT...CAGGCGAGACCGACCGCGCTTCAAT 1956
      |||
      |||
      |||
558 nIleLysAsnPheAspIleLysGlyLeuValThrThrArgValGlns 575
      |||
      |||
      |||
1957 CATTTGAAGAGCGGTGTCAAAATGGAAGTATCCCAAGACAAT 2006
      |||
      |||
      |||
575 erPhe.....GlyGlnThrIlePheGlyGlu 584
      |||
      |||
      |||
2007 CGTGTGGAGACAGATTTGATCGACCGCAATTTAAAGCGAAACTTCC 2056
      |||
      |||
      |||
585 AsnIleGlyAsPlys...SerArgIleGlyValValSerLeuGlnThrG 600
      |||
      |||
      |||
2057 ATATTAGGCGGACAGCGGTGTTCCCGCAATGTTCCCAAGTGAA 2106
      |||
      |||
      |||
600 yTrSerProIaLysSerGlyVal.....ThrPheLysGlyGlyL 615
      |||
      |||
      |||
2107 GGGATTGGCATTTAGCAATCAGCCCAAGAGTTTGGGTGTGCAC 2156
      |||
      |||
      |||
615 yslLysLeuValIle..... 619
      |||
      |||
      |||
2157 GCATCAAGCCACACAATCTGTACAGTTGCGACTGAGCGGTGACAA 2206
      |||
      |||
      |||
620 .....AspG 621
      |||
      |||
      |||
2207 GTTGTCCGAAAA...AACCATTCAGAGATTAAGTATGCTTCA 2250
      |||
      |||
      |||
621 uIleTyHisAlaProIaPAsnTyTrPheAspAlaTrgAsnVal..... 635
      |||
      |||
      |||
2251 TTGAGCAAGCGAGCATGAGGCAATGTACAGCTTGGCGATCAGCGCTCA 2300
      |||
      |||
      |||
636 .....ThrAspValGlnIleAsnLysArgIleLeuPheGlyValTr 649
      |||
      |||
      |||
2301 TTTAATCTC.....ACAAGACTT.....GCCACACTCA 2329
      |||
      |||
      |||
649 ogLysAsnIleAlaGlyLysThrGlyLeuMetPheAsnLeuThrLeuA 666
      |||
      |||
      |||
2330 ACGGCAATCTTACGCA.....GGCGAGACAGCAGCTATACGTTACG 2373
      |||
      |||
      |||
666 snSerAsnIaSerMetAspTyGlyLysAspLeuAspLeuThrIleGln 682
      |||
      |||
      |||
2374 CGCAAGCGCACCCAAAGCGC.....AACCTCAGCCTGTGGGCA 2414
      |||
      |||
      |||
683 GlyHisPheThrAsnAsnGlnGlyThrMetAsnLeuPheValGlnAsp 699
      |||
      |||
      |||
2415 TGGCCAAGACATTTAAT.....CAAGCAGATTAAAGCGCAACA 2455
      |||
      |||
      |||
699 yArgValAlaThrLeuAsnAlaGlyHisGlnIaSerMetIlePheAsnA 716
      |||
      |||
      |||
2456 CATCGGCTTCGACAAATGCTTCAATT.....AAT 2484
      |||
      |||
      |||
716 snLeuValAspSerThrThrGlyPheTrLysProLeuIleLysIleAsn 732
      |||
      |||
      |||
2485 CTAAAGCAACAGCGCTTACAAAGCGCAGCTTGAGCTTCGACAAAGC 2534
      |||
      |||
      |||
733 AsnAlaGlnAsnLeuThrLysAsnLysGlnHisValLeu.....Va 746
      |||
      |||
      |||
2535 TAAGCA..... 2541
      |||
      |||
      |||
746 LysAlaIaArgAsnIleAspTyTrAsnLeuValGlyValGlnGlyAlaSer 763
      |||
      |||
      |||
2542 .....AAGCTAAAGCATTCGACACTCACGCGCAATGTCTCCCTAAGCGAT 2586
      |||
      |||
      |||
```

```

1012 .....LysArgLeuGlnAla 1016
3415 GATAAAGACACCGCTTGGGAAACAGCGGAAACCGGAGCGGCGG 3464
      ||||| ||||| ||||| |||||
1017 LeuLysasp.....GlnArgPheAlaSerLeuGlnSerAl 1028
3465 TACCACCGCCCTTC.....CCCCGGCCCGCGGCGGCGGCGG 3502
      ||||| ||||| ||||| |||||
1028 aalaglValLeuTyrGlnPheAlaProLysTyrGlnLys.ProthAsn 1044
      ||||| ||||| ||||| |||||
3503 ATTGCGCGCAAC.....CGACGCCCGCAACCGCAACCGCAACCG 3540
      ||||| ||||| ||||| |||||
1045 ValTrpAlaSerAlaIleGlyGlyThrSerLeuAsnAsnGlySerAla 1061
3541 CAGCGCGAC..... 3549
1061 aserLeuTyrGlyThrSerAlaGlyValaAlaPheAlaTyrLeuAsnGlyLuv 1078
3550 .....CTGATCAGCCGTTATGCAATAGCGGTTGAGTGAATTTTC 3591
      ||||| ||||| ||||| |||||
1078 algluaAlaIleValaIleGlyGlyPheGlySerTyrGlyTyrSerSerPheSer 1094
3592 GCCACGCTCAACAGC..... 3606
      ||||| ||||| ||||| |||||
1095 AsnGlnAlaAsnSerLeuAsnSerGlyAlaAsnAsnThrAsnPheGlyVal 1111
3607 .....GTTTCGCCCGTACAGAGCAATTTGAC.....C 3634
      ||||| ||||| ||||| |||||
1111 LTrpSerArgLepheAlaAsnGlnHisGlnPheAspPheGlnAlaGln 1128
3635 GCGTGTTCGCCGAAGACCGCGGCAACCGCTTGGACAGCGG..... 3678
      ||||| ||||| ||||| |||||
1128 LylAlaLeuGlySerAspGlnSerSerLeuAsnPheLysSerAlaLeuLeu 1144
3679 .....ATCGGGGACCAACAACTACCGTTGC..... 3705
      ||||| ||||| ||||| |||||
1145 GlnAspLeuAsnGlnSerTyrHisTyrLeuAlaLysSerAlaThrThr 1161
3706 .....CMAGATTTCCGCGCTACCGCGCAACAAACGACGCTGC 3742
      ||||| ||||| ||||| |||||
1161 gAlaSerTyrGlyTyrAspPheAlaPhePheArgAsnAlaLeuValLeu 1178
3743 GCCAAATCGGTATGCAAAAAAACCCTGCGGAGCGGCGGCGGCGGCTGC 3792
      ||||| ||||| ||||| |||||
1178 ysPro.....SerValGlyVal... 1183
3793 TTTTCGACAAACGCGGAGGAAACCTTCGACGACGCGCATC 3842
      ||||| ||||| ||||| |||||
1184 ...SerTyrAsnHisLeuGlySerThrAsnPheLysSerAsnSerAla 1199
3843 GGCACGCGCTTGCACGCGTTCCTTTCGCGCATACGCGATCGGACAGT 3892
      ||||| ||||| ||||| |||||
1199 nValAlaLeuSerAsnGlySer..... 1206
3893 TCACATCGGCATCAGCGCGGCGCGGCTTTCAGTACGCGGACGCTTCA 3942
      ||||| ||||| ||||| |||||
1207 .....SerSerGlnHisLeuPhe 1212
3943 GACGCGATCAGAGCAAAATCCGCGCGGCTGCTGCATTACGCGCATTC 3992
      ||||| ||||| ||||| |||||
1213 AsnAla.....AsnAlaAsnValaIle 1219
3993 GCGAAGATACCGCGAGTTTCGCGGATTCGCGATCGAACCAGACATCG 4042
      ||||| ||||| ||||| |||||
1219 uAlaArgTyr.....TyrTyrG 1225
4043 GCGCAACGCGCTATTTGTCGCAAAACG 4071
      ||||| ||||| ||||| |||||
1225 LysPfnHisTyrPheTyrMetAsnAla 1234
seq_name: SwissProt_40:PMPG_CHLTR
seq_documentation_block:

```

```

ID PMPG_CHLTR STANDARD; PRT: 1013 AA.
AC O84879;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpg precursor (Polymorphic membrane
DE protein G).
DE GN PMPG OR C7871.
OS Chlamydia trachomatis.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -I- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -I- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE001360; AAC68469.1; -
DR InterPro: IPR003368; DUF145.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
DR Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 1013 POTENTIAL.
FT SEQUENCE 1013 AA; 107366 MW; F0927743C0A651DD CRC64;

```

alignment\_scores:

Quality:	199.00	Length:	953
Ratio:	0.431	Gaps:	49
Percent Similarity:	48.478	Percent Identity:	21.301

alignment\_block:

US-09-303-518D-653 x PMPG\_CHLTR ..

Align seg 1/1 to: PMPG\_CHLTR from: 1 to: 1013

```

1615 GGGCATTCGCTTTCGTTCCAGCATTCAGAAATACGATGAAGGGCGAT 1664
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 GlyHisSerLeuThrPheGlnLysnIleArgThrSerThrAsnGlyAlaAl 111
1665 GATTTCGAACACATCAAGACAAAGATCCACCGTTACCATTAACAGCA 1714
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 aleu.....SerAspSerAlaAsnSerGlyLeuPheThrIleGlnGly 126
1715 ATTAAGATATTACTACACCGGCAATTAACACACTGTGATGCAAAAAA 1764
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 helysGlnLeuSerPheSerAsnGlyAsnSerLeuLeuAlaValLeuPro 142
1765 GAATATTCCTACACGCTGTTGGCGAAGAATGCAACCAAAACGAA 1814
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 AlaAlaThrThrAsnAsn.....GlySerGlnThrProThrThrThSe 157
1815 CGGGCGGCTCAATCTGAATTACCAACGGAAGACGCGGATCGCACTTAC 1864
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
157 r..... 157

```

```

1865 TCCTTCCGGCGGACAAATTTAAAGCGCAATATCACGCAAAACAGGCG 1914
158 .....ThiProSerAsnGlyThrIleTyrSerLysThrAsp 169
1915 AAACGTGTTTTCAGCGCGACGACGACACCGCCCTACCAATCATTTAG 1964
170 LeuLeuLeuLeuAsnGlnLysPheSerPhe...TyrSerAsnLeuVal 185
1965 AAGCGGTGTGTCAAAATGGAAGGTATCCCAAGAAGAAATCGTGTGG 2014
185 IserGly.....A 188
2015 ACAACGATTGATCGACCGACATTTAAAGCAAACTTCATTTACG 2064
188 spElGlyAlaIleAsp.....AlaLysSerLeuThrValGln 200
2065 GGC.....GGACAGCGGTGTTCCCGCAATGTCGCAAAATGGAAG 2108
201 GlyIleSerLysLeuLysValPheGlnGlnAsnThrAlaGlnAlaSpG 217
2109 CGAT.....TGGCATTTAAGCAATCACGCGC 2134
217 yGlyAlaCysGlnValValThrSerPheSerAlaMetAlaAsnGlnAlaP 234
2135 AAGCAGTTTTC.....GGTGTGCAACCGCATCAAAAGCCACACATCTGT 2178
234 rolleAlaPheIleAlaAsnValAlaGlyValArgLysGlyIleAla 250
2179 ACAGTTCGAGCTGACGAGGTGTGACAGATGTTACGCAAAAAACCTTAC 2228
251 AlaValGlnAsp.....GlyGlnGlnGlyValSerSerSerThrSerh 265
2229 CGACGATTAAGTATGTTGTTTCATTTAGCAAGACGACATCAGAGGCATG 2278
265 rGlnAspProValValSerPheSerArgAsnThrAlaValGln..... 279
2279 TCAGCTTCCGATCACGCTCATTTAAATCTACAGAGACTTGCACACTC 2328
280 .....Phe 280
2329 AAGCGCAATCTTAGTCAGCGGAGACAGCACATACGTTACGGGCA 2378
281 AspGlyAsnValAlaArgValGlyGlyIleTyrSerTyr..... 264
2379 CGCCACCCAAAGCGCAACCTCAGCTCGTGGCAATGCCCAAGACAT 2428
295 .....GlyAsnValAlaPheLeuAsnAsnGlyLysThrLeuP 307
2429 TT.....AATCAAGCCACA.....TTA 2445
307 heLeuAsnAsnValAlaSerProValTyrIleAlaAlaGlnProThr 323
2446 AAGCGCAACATCGGCTTCGACAAATGCTTCATTTAATCTAAGCAACA 2495
324 AsnGlyGlnAlaSer.....AsnThrSerAspAs 333
2496 CGCCGTCACAAAGCGAGTCTACGCTTCGACAAAGCTTAAGCAACG 2545
333 nTyrGlyAspGlyAlaIlePheCysLysAsnGlyAlaGlnAlaIle 350
2546 TAAGCATTCGCGACTCAAGCGCAATGTCCTTACCGCATTAAGCAATA 2595
350 LysSerAsnAsn.....SerGlySerValSerPheAspGlyGlnGlyVal 364
2596 TTCATTTTGAAGACGCGCTTTACGGAATAATCACGCGGCGAGCA 2645
365 ValPhePheSerSerAsn.....ValAlaAlaGlyLysGln 376
2646 TACGCGATTACCTTAAAGACAGCAATGACGCTGCCGTGGCGACG 2695
376 yGlyValAlaIleTyrAlaLysLysLeuSerValAlaAsnGlyProValG 393

```

---

```

2696 AA.....TTAGCAATTTAAACCTTGACAAAGCCACCATTAACGTAATCC 2742
393 InPheLeuLysAlaIleAlaAsnAspGlyAlaIleTyrLeuGly 409
2743 GCCATTCCACAGATCGCGCAGCGCGCAAAACCGGCACTGGCGCATGC 2792
410 SerGlyGlnLeuSerLeuSerAlaAspTyrGlyAspIleIlePheAspG 426
2793 GCCGCGCGCGCTTCGCGCGCTTATTCGTTACGGCGCAACT 2842
426 yAsnLeuLysArgThrAlaLysGln.....AsnA 436
2843 CGCGAAGATCCCGTTTCAACAGCTGACGCTAAGCGCAATTAAGAGCT 2892
436 laAlaAsp.....ValAsnGlyValThrValSerSerGlnAlaIleSer 450
2893 CAGGAAACATTCGCTTATATGTCGAACTCTTCGGATACCGCAGCGCA 2942
451 MetGlySer.....GlyGlyLys 456
2943 ATTGAACCTGGCGGAAGTCCGAAGGCACTTACACCTG..... 2982
456 sIleThrThrLeuArgAlaLysAlaGlyHisGlnIleLeuPheAsnSpr 473
2983 .....GCTGTCAACATACCGGCAAGCAACCCGTA...AGTCTGACCAA 3024
473 rolleGlnMetAlaAsnGlyAsnAsnGlnProAlaGlnSerSerGlnPro 489
3025 TTGACGATAGTGAAGGAAAGACACACACG.....CTGTCCGAAA 3068
490 LeuLysIleAlaAsnAspGlyGlnGlyTyrThrGlyAspIleValPheAlaAs 506
3069 TCTTAATTTCACCTG...CAAAACGAACAGCTGATCGCGCGCATGGC 3115
506 nGlyAsnSerThrLeuLysThrGlnAsnValThrIleGlnGlnGlyArg... 521
3116 GTTATCAGCTTATCCGAAAGCGGAGATTCGCTGCATTAATCCGCTC 3165
522 .....IleValLeuArgGlnLysAlaLysLeuSerValAsnSer..... 534
3166 AAGCAACAGAGCTTTCGCAACACTGCG.....AAGCGGG 3203
535 .....LeuSerGlnThrGlyLysLeuTyrMetGlnAlaG 547
3204 AGAAACAGAGGCGCGCTTGACGCGCAAAACAGGCAACTT.....GCCG 3247
547 ySerThrLeuAspPheValThrProGlnProProGlnGlnProAlaIa 564
3248 CCNAACACAGGCGGAAAGACACAGCGCA...AGCTTGAAGCGCTG 3294
564 laAsnGlnLeuIleThrLeuSerAsnLeuHisLeuSerLeuSerLeu 580
3295 ATTGCGCGCGGCGCAATGCGACGCAAGCAAGCAAGTTCGCGCAAC 3344
581 LeuAlaAsnAsnAlaValThrAsnProThrProThrAsnProProAlaGlnAs 597
3345 GCGCGCGGACGCA.....GGCGGGAATAATGCCGCGCATTAAGCAGCGG 3388
597 pSerHisProAlaIleIleGlySerThrThrAlaGlySerValThrIleS 614
3389 AG.....GAAGACAAAACGGGTGCAGCGCGATTAAGAC 3423
614 eGlyProIlePhePheGlnAspLeuAspAspThrAlaTyrAspArgTyr 630
3424 ACGGCTTGGGGAAGCGCGCAAGCGGAACCGCGCGGCTACCAACCG 3473
631 AspThrLeuGlySerAsnGlnLysIleAspValLeuLysLeuGlnLeuG 647
3474 CTTCGCCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGCAGACCC 3523
647 yThrGlnProSerAlaAsnAlaProSerAspLeuThrLeuGlyAsnGln 664
3524 AACCGCAACCCCAACGCGACGCGACCTGATTCAGCCGTTATGCCAATAGC 3573

```



```

505 .....ACAGATGCAGAACTGTTGAGATGACGATT 535
114 userSerAlaThrSerThrSerGluTyrAsnProIleIleYsglnArg 131
536 ATATGATGGG.....TGG 549
131 heYsaSpGlyThrPheSerLeuGlyThrLeuValSerGluGlySerLeu 147
550 AAATACCGCTGATTTAAATAAATACCCGTGATTCGTAATCGACAGG 599
148 LysIleHisTyrIleAsnGlnSerGlyAsp..... 157
600 CAGACATATTTGGCGTCTGATGACGACCAACCAATACCGGAAAGTT 649
158 ..SerLysTyrTrp..... 161
650 CATATCATATTGCAAGCGCATATTCTTGCGTGGTGCATACCTTT 699
162 .....ThrPhe 163
700 GCACAAATGATCAGTGTGTCGACAGTCACCTTAGTACGCAAAAT 749
164 ArgArgAsp.....GlyGlyPheThrValAspGlyGly..... 174
750 TAAACATAGCCCATATGCTTTTACCACAGAGAGCTCATTTGGCGACA 799
175 .....GlyLeuGlyValSerGlySerIleThrThrS 186
800 GTGGC.....TCACCAATGTT..... 816
186 erGlyAsnIleAlaIleLeuGlyAsnIleThrSerProGlnIleAsnThr 202
817 .....ATCTAGATGCCCAAAA 833
203 LysAsnIleIleLeuAspThrLysAlaPheGlyGlnTyrAspSerGln.. 218
834 GCAAAAGTGTTAATTAAATGGGTTATTCGAAACAGCAACCCCTATATAG 883
219 .....SerLeuValGlnTyrValTyrProGlyThrG 229
884 GAAAAAGCAATGGCTTCCAGCTAGTTCGTAAGATGGTTCTATGATGAA 933
229 LysGlnGluAsnGlyIleAsnTyrLeuArgLys..... 239
934 ATCTTGCTGGAGATACCACTTCAATTCATTCAGAA.....CCACATGA 977
240 ValArgAlaLysSerGlyGlyThrIleTyrHisGluIleAlaSerAlaG 256
978 AAATGGGAATACTTTTAAAGAC..... 1002
256 nThrGlyLys.....AsnAspGluIleSerTyrThrGlyAsnT 270
1003 .....AAATAATGGCGCA..... 1017
270 hIleuThrThrLysLeuMetGlyLeuArgAsnAspGlyAlaMetValLeu 286
1018 .....GGAATAATC.....GATGCCAAACATAAACA 1043
287 ArgArgSerLeuAlaIleGlyThrIleThrAlaAspGluAsnThrLysAs 303
1044 CTATTCCTACCTTATAGATTAATAAACAGAACCGTTCATGTTTAAG 1093
303 nTyrGlySerProThrProMetGlyGluArgTyrIleAlaLeuGlyAsp. 319
1094 TTTCCTTATCCAGACAGACAGAACCTGTTATCATGCTGACAGGTGG 1143
320 .....AlaAlaThrGly 323
1144 GTCAACAGTTATCGACCACTGATTAATGGAATAATATTTCCCTTAT 1193
324 .....LeuLysTyrIle 327
1194 TGCAAGAGAAAGTGATTAATTAATCTTACGACGACATCAACCAAGGCG 1243

```

```

1 ..... 342
327 eYsgInGlyValTyrAspLeuVal.....GlyAsnThrPsaSerValA 342
1244 CGGGCGGCTTTTATTTTGAAGGTATTTACGGTTCGCGCTTAAACAC 1293
342 La.....SerIleThrProAspSerPhe 349
1294 GAACGTCGCAAGCGCGGCTTCATATCATGATGATGACGATACCGTTAC 1343
350 ArgSerThrArgLysAlaLeuPheGlyArgSerGluAspGlnGlyGly 366
1344 TTGGAAGTAAACGCGCTGGCAACGACCGCTGTCCAAAATCGGCAAG 1393
366 rThrPheMetProGlyThr.....AsnAla 375
1394 GCACGCTGCTGTCACGCCAAAGGGGAAACCAAGGCTCGGTCAAGCTG 1443
375 AlaLeuLeuSerValGlnThrGlnAlaAspValAsn.....AsnAla 388
1444 GCGCAGGTAAAGTCATCTTAGATCAGCAGCGGACGATCAAGCAAAA 1493
389 GlyAspGlyGln..... 392
1494 ACAACCTTTAGTGAATCGCTTGCTCAGCGC..... 1527
393 .....ThrHisIleGlyTyrAsnSerGlyGlyLysMetSerHisT 406
1528 .....AGGGGGCAGGTGCACACTGAATGCCGATTAATCAG..... 1560
406 YrPheArgGlyLysGlyGlnThrAsnIleAsnThrGlnLysGlyMetGlu 422
423 ValAsnProGlyIleLeuLysLeuValThrAspSerAsnAsnValGlnP 439
1561 TTCAAACCCCGAC.....AAACTATTTCCGCTTCCGCGGACGCTTT 1604
1605 GSATTTGAACGGGCAATTCGCTTCCTCCACCGCATTCAAATACCGATG 1654
439 eTyrAlaAsnGlyThrValSerSerIleGlnArgIleLys...PheAspA 455
1655 AAGGGCGCATGATGTCAACCAACCAATCAAGAC.....AAAGATCC 1695
455 snGlyLeuValLeuThrclYalArgProAspGlyIleGlnLeuAspAla 471
1696 ACCGTTACCATTTACAGCAATTAAGATTTACTCAACCGGCAATTAACA 1745
472 ProThrAlaAlaAspGlyThrLysThrIleLeuThrAlaGlyGlyThrAr 488
1746 CAACCTGATACCAAAAAGAAATTCCTACACGCTGCTTGGCGAGA 1795
488 GAlaGlyGlnAsnLysSerTyrValSerIleLysAlaTrpGlyAsn... 503
1796 AAGATGCAACCAAAAAGCAAGCGGCGCTCAATCTGAATTAACCAACGGA 1845
504 .....SerPheAsnAlaSerGlyAsp 510
1846 GAAGCGATCGCACTTTACTGCTTTCCGGCGCAACAAATTTAAACGGCA 1895
511 ArgAlaArgGluThrValPhe.....G 518
1896 TATCAGCAAAACGCGCAACTGTTTTCAGGGCAGACCGACACCGC 1945
518 uValGlyAspGlyGlnGlyPheHisPheTyrSerGlnArgValAlaProA 535
1946 ACGCTACATCATTTAGAGAGCGGTCGTCGTAATAATGAGAGT..... 1989
535 IapThrGlySerThrValGlyProIleGlnLeuArgValAsnGlyGlyLeu 551
1990 ATCCCAAGAGAGAAATCGTGTGGACACGATTCGATCGCACCATTT 2039
552 LeuThrAlaGlySerIleValAlaSerGlySerIleThrThrGluSer.. 567
2040 TAAAGCGAAACTTCATATTCAAGGGCGGCAACGCGGTGCTTCCCGCA 2089

```



```

3604 AGCGTTTCGCGGTACAGGACGAATTGGACCGCGTGTTCGCGAAGACCG 3653
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1091 SerLyS1IleThrPheMetAlaAspHisGlyAspLeuSerAsnSerAsp 1107
3654 CCGGACGCGCGTTTGACAGGCGGATCCGGGACACCAACACTACCGTT 3703
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1107 rTyrtYrProIle.....ValGlyAlaTySerA 1117
3704 CGCAAGATTTCCGCGCTACCGCCACAAACCGACCGCGCAATTCGT 3753
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1117 snTyrcLySerAlaGlyTyraGlnThrPheGlnPheGlyTrpValGly 1133
3754 ATGCAGAAAAACCTCGGACGCGCGCGCTGCGCATCTGTTTTCGCACA 3803
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1134 SerGlyThrThrAlaGly...TrpArgAspGlyIleIleArgIleArg 1149
3804 CCGGACCGGAAACACTTCGACAGGCGATCGGCAATCGGACGCGCTGG 3853
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1149 Y.....AspAsnAlaAsnGlyGlnGlnAlaArgTrpA 1160
3854 CCCACGCGTCCGCTTTTCGGGCAATACGCG..... 3882
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1160 rg.....PheThrMetAspGlyThrLeuAspGlyLys 1172
3883 .....ATCGGCAAGTTTCGACATCGGCGATCGCGCGCGCGC 3917
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1173 ValLeuLeuProGlnThrGlyAlaPheGlyValAsnThrSerAsnGly 1189
3918 GGGTTTATGAGCGGCGGCTTTCACAGCGCATCGAGCGCAAAATCCGCG 3967
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1189 uGlyGlyAsnSerIleThrPheGlyAspSerAspThrGly...IleLysG 1205
3968 GCCGCGTGTGCAATTCGCGCATTCAGCAAGATACCGCGCGGTTTCGCG 4017
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1205 lAsn.....Gly 1207
4018 GGATTTGGCATCGAACCACATCGCGCGACGCGCTATTC...GTCCA 4064
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1208 AspGlyLeuLeuAspIleThrAlaAsnSerValGlnValPheArgPheG 1224
4065 AAAAGCGGAT...TACCGATACGAAACGCAATATCCGCGACCGCGCGC 4111
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1224 nAsnGlyAspLeuTySerTyrIlyAsnIleAsn..... 1235
4112 TTGCGATTCACCGCTAC...CGCGCGCGGCTTAAAGCAGATTATTCATTC 4158
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1236 ..AlaProsnValTyrlleArgSerAspIleArgLeuLysSerAsnPh 1251
4159 AAACGCGCGCACAC 4173
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1252 LysProIleGlnAsn 1256

```

seq\_name: swissprot\_40:AMYH\_YEAST

seq\_documentation\_block:

ID AMYH\_YEAST STANDARD: PRT: 1367 AA.

AC P08640; P08068; PRT: 1367 AA.  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucoan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).  
 GN STA1 OR STA2 OR MAL5 OR YIR019C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 NC NCBI\_TaxID=4932;  
 RX [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Barrer B.G., Badcock K., Bankier A.T., Bowman S., Brown D., Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A., Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,

```

RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RY Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT SPAL";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RY Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPOC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPOC285.13C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MOC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 874
FT CARBOHYD 817 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

```

#### alignment\_scores:

Quality: 199.00 Length: 999  
 Ratio: 0.423 Gaps: 48  
 Percent Similarity: 47.047 Percent Identity: 20.220

#### alignment\_block:

US-09-303-518D-653 x AMYH\_YEAST ..

Align seg 1/1 to: AMYH\_YEAST from: 1 to: 1367

```

1510 ATCGGCTTGGTCACGCGGAGGCGGCGGCGCACTGAATCCGATTAATCA 1559
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
99 lIleGlyValThrGlyProLysGlyThrValGlnIleuTyrllyTrasnG 115
1560 GTTCAACC.....CCGACAACTATATTGGCGCTTC... 1591
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
115 u.AsnThrTyrlleuIleAspAsnProThrAspPheThrAlaThrPheGln 131
1592 .....GCGCGGACGCTTGGATTTGAACGCGG 1617
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
132 ValTyrlaThrGlnAspValaAsnSerCysGlnValTrp..MetProAsn 147

```





```

581 .....ValProthrPro.... 584
3432 GCGGAACAGCGCGAAGCGAACCAGCGCGCTACACCGCGCTTCCCC 3481
585 .....SerSerSerThrThrGluserSerSerAlaProValProthrPro 599
3482 GCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3531
600 SerSerSerThrThrGluser.....SerSerAlaProAlaProthr 613
3532 CCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3581
613 rProSerSerSerThrThrGluserSerSerAlaPro..... 626
3582 TGAATTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3631
626 alThrSerSerThrThrGluserSerSerAlaProValProthr..... 640
3632 ACCGCGTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3681
641 .....ProSerSerSerThrThrGluserSerSerAlaProVal 653
3682 CCGGACACCA...AACACTACGTTGCGAGATTCCGCGCGCTACCGCA 3728
653 lProthrProSerSerSerThrThrGluserSerSerAlaProVal.... 668
3729 ACAAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3778
669 .....ProthrProSerSerSerThrThrGluserSerSerAlaProVal 683
3779 GCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3828
684 ThrSerSerThrThrGluserSerSerAlaProValThrSerSerThr 700
3829 GCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3878
700 rGluserSerSerAla...ProValProthr...ProSerSerSerThr 715
3879 CGGACGCG.....GCAAGTTGCGACATCGCGCA 3904
715 hrGluserSerSerAlaProValProthrProSerSerSerThrThr 731
3905 TCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3954
732 SerSerSerAlaProValProthrProSerSerSerThrThrGluser.. 747
3955 GCGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4004
748 .....SerSerAlaProValThrSerSerThrThrGluserSerSer 762
4005 CGCAGGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4054
762 lAProValProthrProSerSerSerThrThrGluserSerSerAlaPro 778
4055 ATTTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4104
779 ValProthrProSerSerSerThr...ThrGluserSerSerAlaProVal 794
4105 CCGGCGCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4150
794 lProthrProSerSerSerThrThrGluserSerValAlaProValPro 811
4151 .....ATTCAATCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4189
811 hrProSerSerSerSerSerAlaThrSerSerAlaProSerSerThrPro 827
4190 .....ATTGAGCGCTGCTTACCGCGCGCGCGCGCGCGCGCGCG 4221
828 rProSerSerSerThrThrGluserSerSerValProValProthrProSe 844
4222 GCGAAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4271
844 rSerThrThrThrGluserSerSerAlaPro.....ValSerS 856

```

```

4272 CAAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4315
856 rSerThrThrThrGluserSerSerValAlaProValProthrProSerSer 872
4316 .....GTTTCAGCGCTGCTTCCGCGCGCGCGCGCGCGCGCGCG 4342
873 SerAlaThrSerSerSerAlaProSerSerSerThrPro 884
seq_name: SwissProt_40: HXA3_HAEIN

```

```

seq_documentation_block:
ID HXA3_HAEIN STANDARD; PRT; 917 AA.
AC P45355;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme/hemopexin utilization
DE protein A).
GN HXA3.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI N182;
RX MEDLINE=9511556; PubMed=7815944;
RA COPE L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.,
RT "The 100 kDa hem/hemopexin-binding protein of Haemophilus
RL influenzae: structure and localization."
RL Mol. Microbiol. 13:863-873(1994).
CC - FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC - SUBCELLULAR LOCATION: Secreted.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL; U08349; AAA4139.1; -.
KW Transport; Signal; Repeat.
FT CHAIN 1..21
FT SIGNAL 1..21
FT DOMAIN 22..38
FT REPEAT 24..28
FT REPEAT 29..33
FT REPEAT 34..38
FT DOMAIN 111..668
FT REPEAT 111..116
FT REPEAT 111..116
FT REPEAT 203..208
FT REPEAT 203..208
FT REPEAT 277..282
FT REPEAT 399..404
FT REPEAT 629..634
FT REPEAT 663..668
FT DOMAIN 159..170
FT REPEAT 159..164
FT REPEAT 165..170
FT REPEAT 165..170
SQ SEQUENCE 917 AA; 100148 MW; 7FF3BB8C046539D CRC64;

```

```

alignment_scores:
Quality: 198.00 Length: 1129
Ratio: 0.415 Gaps: 55
Percent Similarity: 42.250 Percent Identity: 19.309

```

```

alignment_block:
US-09-303-518D-653 x HXA3_HAEIN ..

```

Align seg 1/1 to: HXA3\_HAEIN from: 1 to: 917

```

442 GAGGAGACTAACGGCATTATGGCGG.....GATTATCATATGCC 485
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 AlasethrInglIeuProInglIeuProInglIeuProInglIeuPro 37
486 GCGTTGGCAAAATTTGTCACAGATGACAGAACCTGTGAGATGACCACTT 535
|||:|||||:|||||:|||||:|||||:|||||:|||||:
37 oGIngluAspLysValValIcIgluAlaIthrPheAspLysThr.... 52
536 AATATGATGGGTGGAAATACGCTGAT.....TTAAATTAATACCT 576
|||:|||||:|||||:|||||:|||||:|||||:|||||:
53 .....IleAlaAspLysMetThrIleAsnInIthrSer 63
577 GATCGTTGCAATGCGACAGCAGACAAATATGCGGCTCGATGANGA 626
|||:|||||:|||||:|||||:|||||:|||||:|||||:
64 AspLysValGlnIleAsn.....TrrPheSer...PheAs 74
627 CGAACCCCAATAACCGCGAAAGTTCAATATTCAGC..... 666
|||:|||||:|||||:|||||:|||||:|||||:|||||:
74 pIleGlyInaAsnLysGluValGluPheLysGlnProSerGluAsnSerV 91
667 ..GCATATTCCTGGCTCGCGTGGCAATACCTTTGCCAAATGCA... 711
|||:|||||:|||||:|||||:|||||:|||||:|||||:
91 AlAlaIerYAsnArGValThrGlyGlyAsnAlaSerGlnIleGlnLys 107
712 ...TCAGGTGTGCGACAGTCACTAGTAGCGAAAAAATTAACATAG 758
|||:|||||:|||||:|||||:|||||:|||||:|||||:
108 LeuThrAlaAsnGlyLysValIyIleuAlaAsnPro..... 119
759 CCCATATGCTTTTTCACACAGAGGCTCATTTGGCGACAGTGGCTCAC 808
|||:|||||:|||||:|||||:|||||:|||||:|||||:
119 ..... 119
809 CAATGTTATCTATGATGCCAAAAGCAAAAGTGTAAATATGGGTA 858
|||:|||||:|||||:|||||:|||||:|||||:|||||:
120 .....AsnGlyVal 122
859 TTGCAACAGCGAACCCCTATATAGAAAAAGCATGCTTCACAGCTAGT 908
|||:|||||:|||||:|||||:|||||:|||||:|||||:
123 IleIleThrLysGlyAlaGluIleAsnValAla...GlyLeuLeuAlaTh 138
909 TCGTAAAGATTGGTCTATGATGAATTTTGGCTGGAGATACCCATTGAG 958
|||:|||||:|||||:|||||:|||||:|||||:|||||:
138 rThrLysAsp.....LeuGluGlnIle..... 145
959 TATCTACGAACACACATCAAAATGSGAAATACTTTTACGACATATAT 1008
|||:|||||:|||||:|||||:|||||:|||||:|||||:
146 .....SerGluAsnSer 149
1009 AATGCGCAGAGAAATGATGATCCAAACATAAACATATTCCTACCTTA 1058
|||:|||||:|||||:|||||:|||||:|||||:|||||:
150 AsnSer.....TyrGlnPheThrAr 156
1059 TAGATTAATAACACGACCGTT.....CAATGTTTATATGTTT 1096
|||:|||||:|||||:|||||:|||||:|||||:|||||:
156 gLysThrLysAspArgGlnValValLysGluGlnValIleAsn.... 171
1097 CTTATCCGAGACACAGACAGAACCTGTTTATCATGCTGCAGGTGGGTC 1146
|||:|||||:|||||:|||||:|||||:|||||:|||||:
172 .....GluGlyGluIle 175
1147 AACAGTTATCGACACAGATGAAATATGAGAAAAATATTTCTTATATGA 1196
|||:|||||:|||||:|||||:|||||:|||||:|||||:
176 LysAlaLysAspPheValValLeuAsnGlyLysGluVal.....IleAs 190
1197 CAAGAAGAAAGGTGAATGATCTTACAGACATCAACAGCGCGG 1246
|||:|||||:|||||:|||||:|||||:|||||:|||||:
190 nLysGlyAsnIleAsnValGluLysAsnSerThrIleAsn.....G 204
1247 GCGGTTGATTTTGGAGGT.....AATTTACGCTCTCCCTTAAAAAC 1290
|||:|||||:|||||:|||||:|||||:|||||:|||||:
204 LysValIyIleuSerSerGlyTyrAsnPheThrPheThrLeuSerAsp 220

```

```

1291 AACCAAAAGTCGCAAGCGCGGCTTCATATCAGTATGCACTACCT 1340
|||:|||||:|||||:|||||:|||||:|||||:|||||:
221 Ser.....GlyIleSerValAlaLeuGlnAspAsnThr.... 231
1341 TACTTGGAAAGTAAACGCG...GTGCGCAACGACCGGCTGTCCAAATCG 1387
|||:|||||:|||||:|||||:|||||:|||||:|||||:
232 .....ValArgGlyIleValGlnAsnGluGlySerIleLysAlaG 245
1388 GCAAGGACGCGTCTGCTTCAGCCCAAGGGGAAAAACCAAGGCTCGTC 1437
|||:|||||:|||||:|||||:|||||:|||||:|||||:
245 LysGluIleThrLeuSerAlaLysGlyArgLysGluAlaLeuAspSerLeu 261
1438 AGCGTGGCGCGGTAAAGTCATCTTATGATCAGACGCGGACATCAAG 1487
|||:|||||:|||||:|||||:|||||:|||||:|||||:
262 ValMetAsnAsnGlyValLeuGluAlaThrLysValSerAsnArgAsnG 278
1488 CAAAAAACAGCCCTTATGTAATCGGCTTGTCAGCGCGGCGGACG 1537
|||:|||||:|||||:|||||:|||||:|||||:|||||:
278 Lys.....ValValLeuSerAlaAspAsnV 287
1538 TGCACGTAATCCGATATACGTTCAACCCGACAAACTCTATTTGCGC 1587
|||:|||||:|||||:|||||:|||||:|||||:|||||:
287 AlGlnLeuAsnAsnGlnSerAsnIleLysGlyIleValAsnPheGly 303
1587 ..... 1587
304 ThrGluValIthrSerAsnGluAspLysLysLeuLysIleThrSerLys 320
1588 .....TTTCGCGCGGACGTTTGG 1606
|||:|||||:|||||:|||||:|||||:|||||:|||||:
320 rGlySerLysValIthrSerProLysIleAsnPheLysGlyLysSerVal 337
1607 ATTGAACGG.....CATTCGCTTTCG 1629
|||:|||||:|||||:|||||:|||||:|||||:|||||:
337 snIleAsnGlyAsnPheGlyArgGluAsnSerGlyThrIleIstYAsnGlu 353
1630 TTCACCGCATTCCAATACCGAT.....GAAGGGCGAT 1664
|||:|||||:|||||:|||||:|||||:|||||:|||||:
354 GluArgLysThrLeuAsnThrGluValAsnIleAspValProGlyAlaG 370
1665 GATTGTCACCCACATCAAGACAAAGATCCACCGTTACC..... 1704
|||:|||||:|||||:|||||:|||||:|||||:|||||:
370 uAsnIleArgIleAlaAspAspLysAspAsnThrGluThrAspSerPhe 387
1705 .....ATTACAGCAATTAAGTATTACT 1728
|||:|||||:|||||:|||||:|||||:|||||:|||||:
387 leGlnThrGluAlaLeuSerSerLeuLeuAlaAsnAsnGlyLysValAsn 403
1729 ACAACGCAATTAACAACACACTGGATAGCAAAAAAGAAATTCCTACAA 1778
|||:|||||:|||||:|||||:|||||:|||||:|||||:
404 LeuLysGlyAsnAspValAsnIleSerGly.....AsnIleAsnIleAs 418
1779 CGGTGGTGTGGCGAGAAAGATCAACCAAA...ACGAC...GGCGCGC 1822
|||:|||||:|||||:|||||:|||||:|||||:|||||:
418 pSerPheArgGlyThrAspSerLeuLeuLysLeuThrAsnGlnGlyHis 435
1823 TCAATCTGAATTACCAACCGGAAGACGATGCGACTTTACTG..... 1866
|||:|||||:|||||:|||||:|||||:|||||:|||||:
435 IeAsnIleAspHisAlaAspIleAsnSerLysGlyArgLeuPheValAl 451
1867 .....CTTCGCGCGGACAAATTTAAACGCAATATTCGCAACAAA 1910
|||:|||||:|||||:|||||:|||||:|||||:|||||:
452 ThrSerLeuGlnAsnAspValAspPheLysSerAsnIleThrIleThrAs 468
1911 CGGCAAACTGTTTTCAGCGGACAGCCGACACCGCTACATCATAT 1960
|||:|||||:|||||:|||||:|||||:|||||:|||||:
468 pSerLysIle.....AsnL 473
1961 TAGGAAGCGGCTGCAAAAATGGAAGGTATCCCAAGAGAAATCGTG 2010
|||:|||||:|||||:|||||:|||||:|||||:|||||:
473 euGlyAsnGly.....AlaMetLysLeuGlyArgSerVal 484

```

```

2011 TGGACACGATTCGATCGACCGACATTTAAAGCGAA..... 2049
      :::::|||||
485 AsnGluAspSerPyrAlaThrArgTrpLysArgAlaGluLysSerGlnAr 501
2049 ..... 2049
501 gLysLysPheAsnValAspMetArgAsnValAlaPheAsnGluValAspG 518
2049 ..... 2049
518 luValIleLeuAlaGlyGlyPheGluLysValAsnLeuAspLysIleVal 534
2050 ..... AACTTCATATTACGGCGGACAAAGCGGTGTTTC 2084
      |||||:::|||||
535 AlaThrGlyGlnThrAsnPheryIleAspGlyGly.....ValSe 548
2085 CGCGAATGTTCCCAAAGTGAAAGCGATTGGCATTTAAACAATCACGCC 2134
      |||||:::|||||
548 rArgAsnAsnSerArgTrpGlu..... 555
2135 AACGAGTTTTCGTTGCGACCGCATCAAGCCACATCTGTACA... 2181
      :::::|||||
556 .....TyrGlyValIleuAspLeuAspPTrArgThrGlnLeuSerGlu 569
2182 .....CGTTCGGAAGTGG..... 2193
570 LeuGluGlnGlyArgArgArgTrpArgTrpTyrArgAspLeuAspLeuAs 586
2194 .....A 2194
586 pMetAsnLysAlaTyrLeuTyrArgPheAspLeuPheAlaLysAsnAsnS 603
2195 CGGGTCTGCAAGTGT.....ACCGAAAAACCATTAACCGAC... 2232
      :::::|||||
603 erGlyArgSerThrIleLysGlyThrGluIleAsnIleSerAsnSerAsn 619
2233 .....GATTAAGTGAATTC 2246
620 IleAsnLeuLysAsnGlyPheValIleLeuAlaGluAsnIleLysIle 636
2247 TTCATGAGCAAGACCGACATC.....AGAGCAATGTCA 2281
      :::::|||||
636 uAspAsnSerLysValAspIleThrPheAspLysAspAsnSerGlnAspT 653
2282 GCCTTCGCGATCAGCTCATTTAAATCTCACAAGA...CTTGCCACATTC 2328
      |||||:::|||||
653 hIreuAlaGlnThrAsnArgLeuGlyMetAsnGlyLysValSerMetIle 669
2329 AACGCAATCTTAGTCAGGCGGAGACAGCACTATACGTTACGGCGCA 2378
      |||||:::|||||
670 AsnSerHisIleLysIleValGlyAspArgLysGluGlyIleSerProth 686
2379 CGCGACCAAAAGCGCAACCTC.....AGCCTCGTGGCGCAATGCC 2419
686 rGlyThrTyrAlaThrMetPheLeuIleGlyIleLeuIleGlyGluLys 703
2420 AACCAACATTTAAATCAAGC.....ACATTTAAAGCGC 2451
      :::::|||||
703 erSerIlePheValLysSerHisGlnGlyTyrThrPheLysThrAspGly 719
2452 AACACATCG...GCTTCGACAAATGCTTCATTTAATCAACAACAGC 2498
      :::::|||||
720 AspThrLysIleAlaGlyLysAsnSerLysGluAspLeuLysIleThrAl 736
2499 CGTACAA.....AACGGAGTGC 2515
736 alIeAsnThrGlyGlyArgAlaAlaGluGluValIleuIleAsnGlyAlaL 753
2516 TGACGCTTTCGACAGCAAGCTAAGCAAGCAATTCGCACTCAAC 2565
      |||||:::|||||
753 euGlySerIleAspAsn...AspAlaAsnIleAlaAsnMetAlaPheThr 768
2566 GGCAATGTCTCCTAGCCGATTAAGCAGTATTCATTTTGAAGAACAGCG 2615

```

```

769 IlegLysAspSerAlaAsnThrLysThr.....IlegLysAsnAlaAs 783
      :::::|||||
2616 CTTTACCGCGAAAAATACAGCGCGCAAGATACGGCATTTACACTTAAAG 2665
      :::::|||||
783 pIleThrAlaLeuAlaProAsnGlyGlyThrAlaTyrLysSerSerLysA 800
2666 ACAGGAAATGAGCGCTGCGGTGCGGCAAGCAATTAAGCAATTTAACTT 2715
      |||||
800 spValGlu..... 802
2716 GACAAGCGCACCATTAACATCAATTCGCTATCGACAGATGCGGAGG 2765
      :::::|||||
803 .....IlegValLysProAsnSerAspPhe..... 811
2766 CGCGAAACCGGACGTGCGGAGATGGCGCGCGCGCGCTTCGCGCGGT 2815
      :::::|||||
812 .....ThrPheGluLeuProArg..... 818
2816 CCTATTATTCGTTACGCCGCCCACTTCGGCAGAAATCCGTTCAACACG 2865
      |||||:::|||||
819 .....GlyLysAsnLeuAsnGln 824
2866 CTGACGGTAAACGCG.....AAATGAACGCTCAGGGAACATTCG 2906
      :::::|||||
825 ThrLysIleAsnGlyAlaSerThrLysLeuSerGluArgGlyPheAlaAr 841
2907 CTTTATGCGGAACTCTTCGCTACCGCAGCGGCAATTAAGCTGCGCG 2956
      :::::|||||
841 gLeuItrAspLysIleAsnGlyValArgAlaSerAsnLeu..... 854
2957 AAAGTTCGGAAGCATTACACTTGGCTGTCAACAATACCGGCACAGAA 3006
854 ..... 854
3007 CCCGTAAGTCTCGACGAATTAAGCGTGTGAGAGAAAGAC..... 3048
      |||||:::|||||
855 .....SerAlaGluGlnLeuAsnValThrAspAlaSerGluLysIleI 869
3049 AACACACCGCTGTCGCGAAATCTTAATTC.....ACCTGCAAAACG 3091
      |||||:::|||||
869 eAsnThrLysLeuValSerSerLeuAspValGluLysLeuValSerValA 886
3092 AACACGTCGATCGCGCGCATGGCTTATCCGCAAGACGCGC 3141
      |||||:::|||||
886 lAlaLysAspAlaGly..... 891
3142 GAGTTCGCGCTGCATTAATCGGTCAAAGACAAAGCTTCCGCAAACT 3191
      |||||:::|||||
892 .....AsnGlyCysGluGlnGlnGlnPheGlyAspLysG 903
3192 CGGC.....AAGCGGAGAAACAGAGCGC 3216
      :::::|||||
903 yAsnAsnThrLysValSerValGlyGluLeuGluAla 915
seq_name: SwissProt_40:MAP4_MOUSE
seq_documentation_block:
ID MAP4_MOUSE STANDARD: PRT: 1125 AA.
AC P27546;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 4 (MAP 4).
GN MAP4 OR MAP4
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042100; PubMed=1718985;
RA West R.R., Tendarge K.M., Olmsted J.B.;

```



```

2830 AGCGCCGCACTTCGGAGAAATCCGTTTCAACAGCCTGACGGTAAACGG 2879
    ||| : : : : : ||| : : : : :
416 ThrGluValValLeuProSerSerProIleThrThrLeuThrLysAspVa 432
2880 CAAATTGAACGGTCAAGGGAACATCCGTTTATGTGGAACCTCTTGCGCT 2929
    ||| : : : : :
432 ThrLeu.....ProLeuGluAlaG 439
2930 ACCGAGCGCGCAATTTGAAGCTGGGGAAGTTCCGAAGGCACCTAACCC 2979
    ||| : : : : : ||| : : : : :
439 IuArgProLeuValThrAspMetThrProSerLeuGluThrGluMetThr 455
2980 TTGGGTGTCAACATACCGGCAACACCGCTTAAGTCTGAGCAATTAAGC 3029
    ||| : : : : : ||| : : : : :
456 Leu.....GlyLysGluThrAlaProProThrGluThrAs 467
3030 GGTAAGTGAAGAAAGACACACACCGCTGTCCGAATTTAAATTCA 3079
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
467 nLeuGluMetAlaLysAspMetSerProLeuProGluSer...GluValT 483
3080 CCTGCAAAACGACACGTCGATGGCGGCGCATGGCGTTATCAGCTTATC 3129
    |||| : : : : : |||
483 nLeuGluLysAspValValIleLeuProGlu..... 493
3130 CGCAAGAGCGGAGCTTCGCTGCATATCCGGTCAAGAACAGACAGCT 3179
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
494 ThrLysValAlaGluPheAsnValThrProLeuSerGluGluGluVal 510
3180 TTCGCAC.....AACTGGCAAGCGGGAAGAAACAGAGCGCGCTTGA 3223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
510 ThrSerValLysAspMetSerProSerAlaGluThrGluAlaProLeu 526
3224 CGCAAAACGAGCACACTTGCCTGCAAAACAGAGCGGGAAGAAACAGAC 3273
    |||| : : : : : ||| : : : : :
527 ..AlaLysnAlaLysPheLeuHisSerGlyThrGluLeuLeuValAsp 542
3274 GCGCAAGC.....CTTGACGCGCTGATTCG 3299
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 SerMetAlaProAlaSerAspLeuAlaLeuProLeuGluThrLysValAl 559
3300 GAGC.....GGCGCATGGCCACGAAAGAGCGAAGATGTTG 3337
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
559 aThrValProIleLysAspLysGlyThrValGluThrGluGluLysPro 576
3338 CGAAGCGCGCGGAGCGAGCGGGAAGTCCGCGC..... 3375
    ||| : : : : : ||| : : : : :
576 rgLysPheArgLeuAlaSerMetGlnHisLysGlyGlnSerThrVal 592
3376 .....ATTATGCAAGCGGAGGAGAGAGAAAA 3401
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
593 ProProCysThrAlaSerProGluProValLysAlaAlaGluGluMetSe 609
3402 AGGGGTGAGCGGATTAAGACACCGCTTGGCAAGACGCGGAGAGCG 3451
    : : : : : ||| : : : : : ||| : : : : :
609 rThrLeuProIleAspAlaProSerProLeuGluLysnLeuGluGlnLysG 626
3452 AAACCGCGCGGCTACACCGCTTCCCGCGCGCGCGC..... 3490
    |||| : : : : : ||| : : : : :
626 IuThrProGlySer.GlnProSerGluProCysSerGlyValSerArgG 642
3491 .....GCGCGCGCGGATTTGCCGCAAC.....CGCA 3518
    ||| : : : : : ||| : : : : :
642 nGluGluAlaLysAlaAlaValGlyValThrGluLysnAspIleThrThrP 659
3519 GCGCCAGCGCAACCGCAACGAGCGGACCTGATCAGCGCTTATGCCA 3568
    |||| : : : : : ||| : : : : :
659 roProCysnLysGluProProProSerProGluLysLysAla...LysPro 674
3569 ATAGCGTTTGAAGTATTTCCGCCAGCTCAGACGCTTTTCCCGCTA 3618
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
675 LeuAla.....ThrThrGlnProAlaLysThrSerThrSerLysAl 688
3619 CAGAGCAATTTGACCGCGTGTGGCGGAAGACCGCGCA..... 3658

```

```

688 aLysThrGlnProThrSer...LeuProLysGlnProAlaProThrThrS 704
3659 .....ACGCCGTTTGACACAGCGCATCCGGAGACCA 3691
    ||| : : : : : ||| : : : : :
704 eArgLysLeuAsnLysLysProMetSerLeuAlaSerGlySerValPro 720
3692 AACACTACCGTTGCAAGATTTCCGCGCTACCGCAAC..... 3730
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
721 AlaAlaProHisLysArgProAlaAlaAlaThrAlaThrAlaArgProse 737
3731 .....AACCG...ACCTGGCGCAAT 3749
    |||| : : : : : ||| : : : : :
737 rThrLeuProAlaArgAspValLysProLysProIleThrGluAlaLysV 754
3750 CGGTATGCAGAAAACCTCGCAGCGGCGCGCTGCGATCCTGTTTCG 3799
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
754 aAlaGluLysArgThrSerProSerLysProSerSerAlaProAlaLeu 770
3800 ACAACCGAGCGGCAACACCTTCGACAGCGCATCGCAACTCGGACGG 3849
    |||| : : : : : ||| : : : : :
771 LysProGluProLysThrThrProThrValSerLysAlaThr..... 784
3850 CTTGCCACGCGTCCGTTTTCGGCAATACGCGATCGCAGTTTCACAT 3899
    ||| : : : : : ||| : : : : :
785 .....SerProSerThrLeuValSerThrGlyProS 795
3900 CGGCATCAGCGCGGCGGCTTTTATAGTCGCGACGCTTACAGCGCA 3949
    |||| : : : : : ||| : : : : :
795 eSerArgSerProAlaThrThrLeuProLysArgProThrSerThrLys 811
3950 TCAGAGGCAAAATCCGCGCGCGTGTGCATACGGCATTCAGGCAAGA 3999
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
812 ThrGluLysProAlaAspValLysArgMetThrAla..... 824
4000 TACCGCGAGTTTCCGCGGAT...TCGCGATCAGACCGACATCGGCGC 4046
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
825 ...LysSerAlaSerAlaAspLeuSerArgSerLysThrThrSerAla 839
4047 AACCGCTATTGTGCCAAAAGCGGATTCACGATACGAAAGATCAATA 4096
    |||| : : : : : ||| : : : : :
840 .....SerSerValLysArgAsnThr...ThrProThrGlyAla 851
4097 TCGCCACCGCGGCGCTTGATTCACCGCTACCGCGCGGCGCATTAAGCA 4146
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
852 AlaProProAlaGlyMetThrSerThrArgValLys..... 863
4147 GATTATTCATTCAAAACGCGGCAACACATTTCCATCAGCGCTTATTGAG 4196
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
863 ..... 863
4197 CCTGTCTATACCGATCCGCTTCCGCAAAAGTCCGAACCGCG...TCA 4243
    |||| : : : : : ||| : : : : :
864 .....PrometSerAlaProSerArgSerSerGlyAlaLeuSer 876
4244 ATACCGCGGATTTGCGCGAGGATTTCCGCAAAACCGCGACTCGCGAATG 4293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
877 ValAspLys.....LysProThrSerThrLysProSer..... 887
4294 GCGCTAAACGCGCAATCAAGGTTTCAGCGCTGCTCCGCTGCGCGC 4343
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
888 .....SerSerAlaProArgValSerArgLeuAlaThrThrValSerA 902
4344 CGCA.....AGGCGCGCAATTTGAAGCGACAGCAGCAGCG 4378
    ||| : : : : : ||| : : : : :
902 IarProAspLeuLysSerValArgSerLys.ValGlySerThrGluAsnI 918
4379 CGGGCATCAA 4388
    |||| : : : : :
918 eLysHisGln 921
seq_name: SwissProt_40:TCNA_TRYC

```



```

1311 GGGCGTTCATATGAGTGTGGCAGTACCGTACTTGAAGTAACGCGG 1360
      |||
274 .....LeuTrpLeuThrAsp... 278
      |||
1361 TGGCAAGACGCGCTGTCCAAATCGCAAAAGCAGCCGTGCTGTTCA 1410
      |||
279 .....AsnGlnArgIle..... 282
1411 GCCAAGGAGAAACCAAGCGCTCGTGCAGCGTGGGCGAGCGTAAGTCAT 1460
      |||
283 .....TyrAsnValGlyGlnValSerIleGlyAsp..... 292
1461 CTTCAGATCAGCAGCGGAGCATCAAGCAAAACCAAGCTTTAGTGAAA 1510
      |||
293 .....GluAsnSerAlaTyrSerSerV 300
1511 TCGGGCTGTGTCAGCGGAGGCGGAGCGGTGCAACTGAATGCCGATATGAG 1560
      |||
300 al.....Leu 301
1561 TTCAACCCCGACAACTCTATTTCGGCTTCGCGGCGAGCATTTGGATT 1610
      |||
302 TyrLysAspAspLysLeuTyr..... 308
1611 GACGCGCATTCGCTTCGCTCCACCGCATTCAAAATACCGATGAA... 1656
      |||
309 .....CysLeuHisGluIle...AsnThrAspGluValT 319
1657 .....GGGCGATGATTCACACGAC..... 1677
319 yTrSerLeuValPheAlaArgLeuValGlyGlnLeuArgIleLysSer 335
1678 .....AATCAAGACAAGAAATCCACGTTACCATTTACAGCAATA...A 1718
      |||
336 ValLeuArgSerTrpLysAsnTrpThrAlaThr.CysProAlaPheAlaP 352
1719 AGATTTACTACACCGCGCATTAACACAACTTGGATAGCAAAAAAGAAA 1768
      |||
352 rGlnLeuIleGlnProLeuArgArgGlnValValValValProLeu 368
1769 TTGCGTACACAGGTTGGTTGGCG.....AGAAGATGCAACCAACG 1812
      |||
369 SerProAlaGlnValLeuValLeuAlaPheCysArgGlnArgLeuProLysAr 385
1813 AACGGGCGGCTCAATCTGATATACCAACCGCAAGGAGCGATCGCATTT 1862
      |||
385 gMetGlyGlySerTyrArgCysValAsn..... 394
1863 ACTGCTTTCGGCGAGACAAATTTAAACGCGCAATATCAAGCAACAAAG 1912
      |||
395 .....AlaSerThr 397
1913 GCAACGTGTTTTTCAGCGGAGACCGGACCGCAGCGCT...ACAATCAT 1959
      |||
398 AlaAsn.....AlaGlnArgValArgAsnGlyLeuLysPheAl 410
1960 TTAGGAAGCGGCTGTCAAAAATGGAAGTATCCCAAGAGGAATCGT 2009
      |||
410 agIyValGlyGlyAlaLeuTrpProValSerGlnGlnGlyAsnang 427
2010 GTGGGACAACGATTGGATCGACCGCAAT...TTAAACGCAAAACTTCC 2056
      |||
427 InArgTyrArgPheAlaAsnHisAlaPheThrLeuValAlaSerValThr 443
2057 ATATTGAGGCGGAGCAAGCGGTGTTCCCGCAATGTT.GCCAAAGTGA 2105
      |||
444 IleHisGlnAlaProArgAlaAlaSerProLeuLeuGlyAlaSerLeuAs 460
2106 AGGCGATTGGCATTTAAGCAATACGCCCAACAGCTTTGGTGTGCGAC 2155
      |||
460 p.....SerSerGlyGlyLysLysLeuLeuGlyLeuSerT 472
2156 CGCATCAAAAGCCACAAATCTGTACACGTTGGAGTGGAGCGGTGTGACA 2205
      |||
472 yTrAspGlnLysHis.....GlnTrpGlnProIleTyr 482
2206 AGTTGTACGCAAAAAACCATTACCGAGCATTAAGGTATGTTTCATAG 2255
      |||
483 GlySerThrProValThrProThr..... 490
2256 CAAGACCGACATCAGAGGCAATGTCAAGCTTCCGATCAACGCTATTAA 2305
      |||
491 .....GlySerTrpGlnGlyLysArgTyrHisLeuVal 502
2306 ATCTCACA.....GGACTTCCCACTCAACGCAATCTT 2340
      |||
502 alLeuThrMetAlaAsnLysIleGlySerValTyrIleAspGlyGlnLeu 518
2341 AGTCAGCGGAGACACCGCACTATACGGTTACGCGAAGCCACCAAAA 2390
      |||
519 LeuGlnGly..... 521
2391 CGGCAACCTCAGCCTCGTGGCAATGCCAAGCAACATTTAATCAAGCA 2440
      |||
521 ..... 521
2441 CATTAACGCGCAACATCGGCTTCGAGCAATGCTTCATTATATCAAG 2490
      |||
522 .....SerGlyGlnThrValAlaProAspGlyArgThrProAspIleSer 536
2491 AACACGCCCTACAAAAGCGGAGCTGACGCTTCGCGAAGCGTAAGGC 2540
      |||
537 HisPheArgVal.....GlyGlyTyrLysArgSerAsp.....MetPr 549
2541 AACGTAAAGCATTCGCGACTCAACGCGCAATGTCTCCCTAAGCGGATAG 2590
      |||
549 oThrIleSerHisValThrValAsn...AsnValLeuLeuTyrAsnArg 565
2591 CAGATTTCATTTGAAACACGCGCTTTACCGGAAAAATCACGCGCGG 2640
      |||
565 rGlnLeuAsnThrGlnGluIleArg..... 573
2641 AAGGATACGCGATTACCTTAAAGACAGCGAATGAGCGTGGCGCTCGGG 2690
      |||
574 .....ThrLeuPheLeuSerGlnAsp.....LeuIleG 583
2691 CACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTTACCTGAAT 2740
      |||
583 yThrGln.....AlaHisMetAspSerSerAspSerAlaHis 597
2741 CCGCCTATCGACACGATGCGGAGCGCGCAACCGGAGAGTGGGAGAT 2790
      |||
597 lSerThrProSerThrProAlaAspSerSerAlaHisSerThrProSer 613
2791 GCGCGCGCGCGCTTCGCGCGCTCCCTATATTCGCTTACGCGCGCAAC 2840
      |||
614 ThrProValAspSerSerAlaHisSer.....ThrProSerThr 626
2841 TTGGGCGAAG...TCCCGTTTCAACAGCTGACGAGTAACGCAATTTGA 2887
      |||
626 rProAlaAspSerSerAlaHisGlyThrProSerThrProValAspSerS 643
2888 ACGGTACGAGACATTCGCTTATGTCGGAACCTTGGGCTACCGCAG 2937
      |||
643 erAlaHisGlyThrProSerThrProAlaAspSerSerAlaHisGlyThr 659
2938 GGCMAATGGAAGCTGGGGAAGTTCCGAGGCACTTACACTTGGCTGT 2987
      |||
660 ProSerThrProValAspSerSerAlaHisSerThrProSerThrProVa 676
2988 CAACATATCCGCGACAGACCCGTAAGTCTCGAGCAATTCAGCGTGTG 3037
      |||
676 lAspSerSerAlaHisSerThrProSerThrProValAspSerSerAlaHis 693
3038 AAGGAAAGACACACACCGCTGTGCGAAATCTTAATTTCACCCGCA 3087
      |||

```



```

693 IsGlyAlaProSerThrProAlaAspSerSerAlaHisGlyThr.Prose 709
3088 AACGACACGTCGATCGCGGCGATGCGTTACGTTATCGCGAAAGA 3137
709 rThrProValAspSerSerAlaHisGlyThrProSerThrProAlaAspS 726
3138 CGCGGAGTTCCGCGTCATATCCGGTCAAAAGAACAGACTTTCGACA 3187
726 eSerAlaHisSer.....ThrProSerThrProAla 736
3188 AATCGCGCAAGCGGAGAAACAGAGCCGCTTACGCGCAAAACAGGCA 3237
737 AspSerSerAlaHisSerThrProSerThrProAlaAspSerSerAlaH 753
3238 CAACCTGCGG.....CCAAACACAGCGCGGAAAGAACACAGCGCGAAG 3281
753 sSerThrProSerThrProValAspSerSerAlaHisGlyThr..... 767
3282 CCTTGACGCGCTGATTCGCGCGCGGCAATGCCACCGGAAAGGAGAAA 3331
768 .....ProSerThrProAlaAspSerSerAlaHis 777
3332 GTGTTGCGCAACCGCGCGCGGACAGCGCGGAAATGCCGCGATTATG 3381
778 SerThrProSerThrProAla..... 784
3382 CAGGCGGAGAGAGAAAGAACGCGGTGACGCGGATTAAGACACGCGCTT 3431
785 .....AspSerSerAlaHisGlyThrProSerThrProValAspSer 796
3432 GCGGAAACAGCGGAAAGCGGAAACCGCGCGCTACACCGCGCTTCCGCC 3481
796 AlaAspSerSerAlaHisSerThrProSer...ThrProValAspSer 811
3482 GCG.....CCGCGCGCGCGCGCGGATTTGCGCGAACCGGACGCC 3522
812 AlaHisGlyThrProSerThrProValAspSerSerAlaHisSerThrPr 828
3523 CAACGCGCAACCCCAACCGCGCGCGACCTGATCAGCGCTTATCCGCAATG 3572
828 oSerThrProValAspSerSerAlaHisGlyThrProSerThrProVal. 844
3573 CGGTTGAGTGAATTTCCGCCACGCTCAACAGCTTTTCCGCGGTACAG 3622
845 .....AspSerSerAlaHisSerThrProSer 853
3623 ACGAATTGACCGCGGTGTTGCCGAAGCGCGCGACGCGCTTGAGCA 3672
854 ThrProValAspSerSerAlaHisSerThrProSerThrProAlaAspSe 870
3673 ACGCGCATCGGAGACACCAACACTACCGTTGCAAGATTTCGCGCGCTA 3722
870 rSerAlaHisGlyThrProSerThrProValAspSerSerAlaHisSer 887
3723 CGGCGCAACAAACGACCTCGCCAAATCGGTTGCAAGAAAACCTCGGCA 3772
887 hrProSerThrPro.....AlaAspSerSerAlaHisSerThrProSer 901
3773 GCGGCGCGCTCGCATCTGTTTGCACA..... 3802
902 ThrProValAspSerSerAlaHisSerThrProSerThrProAlaAspSe 918
3803 .ACCGGACCGGAAACACTTCGACAGCGCATCGCAATCGCGACGCGCT 3851
918 rSerAlaHisGlyThrProSerThrProValAspSerSerAlaHisGlyT 935
3852 TGCCACGCGTGGCGTTTGGCGCATACGCGCATCGCGAGTTGCAACATG 3901
935 hrProSerThrProAlaAspSerSerAlaHisSerThrProSerThrPro 951
3902 GCATACGCGCGCGCGGTTTAGTAGCGGACGCTTTCACAGCGCATC 3951
952 AlaAspSerSerAlaHis.....SerThrProSerThrProAlaAspSe 966

```

```

3952 AGAGCAAAATCCGCGCGCGCTGCTGATTCAGCGCATTCAGGCAATA 4001
966 rSerAlaHisSerThrProSerThrProValAspSerSerAlaHisSer 983
4002 CGCGCAGGTTTCGCGGATTCGCGCATGCAACCCACATCGCGCAACGC 4051
983 hrProSerThrProAlaAspSerSerAlaHisSerThrProSerThrPro 999
4052 GCTATTGCTCCAAAGCGGATTAACGATACGAAACGTCATATCGCC 4101
1000 AlaAspSerSerAlaHisSerThrProSerThrProAlaAspSerSer 1015
4102 ACCCGGCGCTTCGATTCACGCGCATCGCGCGGCGCATTAAGGAGATTA 4151
1016 .....AlaHisSerThrProSerThrPro..... 1024
4152 TTCATTCAACCGCGCGCAACATTTCCATCAGCGCTTATTTGAGCCTGT 4201
1024 AlaAspSerSerAlaHisSerThr...ProSer..... 1033
4202 CCTATACCGATGCGCGCTTCGCGCAAGTCGACAGCGCGGTCAATCGCC 4251
1034 .....ThrProAlaAspSerSerAlaHisGlyThrPro.. 1044
4252 GTATTGCGGAGATTTGCGCAAAACCGCGCGTCCGATGCGCGCTAAA 4301
1045 .....SerThrProAlaAspSerSerAlaHisSer 1055
4302 GCGGAAATCAAGGTTTCACGCTGCTCCACGCTGCGG 4342
1055 hrProSerThrProValAspSerSerAlaHisSerThrPro 1068

```

seq\_name: SwissProt\_40:P2P\_LACPA

```

seq_documentation_block:
ID P2P_LACPA STANDARD; PRT; 1902 AA.
AC 002470;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-
DE associated serine proteinase) (LP151).
GN PRP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1597;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 151;
RX MEDLINE=92381481; PubMed=1512565;
RA Holick A., Naes H.;
RT "Cloning, sequencing and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151."
RJ J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
RP SEQUENCE OF 189-196.
RX MEDLINE=9222694; PubMed=1564442;
RA Naes H., Nissen-Meyer J.;
RT "Purification and N-terminal amino acid sequence determination of the
RT cell-wall-bound proteinase from Lactobacillus paracasei subsp.
RT paracasei".
RJ J. Gen. Microbiol. 138:313-318(1992).
RL
CC -I- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -I- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some subsite preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.

```



```

995 helysleuaspserlysalaprophrvalarghisvalaleuSerAla 1011
1108 ACAGCAAGAGAACCTTTATCATCTGAGGTGGGTCAACGATTATCG 1157
1012 LysThr..... 1013
1158 ACCCAGACTGATTAATGGAGAAATATTTCTTTATGCAAGAAAGAAAG 1207
1014 .....GluAsnGlyLysT 1018
1208 GTGATTTGATTACTTACAGCAATCAACCAAGCGCGCGGTTCAT 1257
1018 hclInrTyrrleuthralaglualalaspAspleuSerGlyLeu... 1033
1258 TTTGAGGTAATTTACGGTCTCGCCTAAACAAAGAAAGCTGGCAAG 1307
1034 .....AspAlaThrLysSerValLysThr..... 1041
1308 CGGGGCGTTCAATTCAGTGGCGGTACCGTTACTTGAAAGTAAAGC 1357
1042 .....AlaIleAsnG 1045
1358 GCGTGCAACAGACCGCTGTCCAAATCGCAAGGACGCGTGTGTT 1407
1045 lvalThrAsn.....Leu 1049
1408 CAAGCCAAAGGAGAAACCAAGCGTGGTACGGCGAGGTTAAAT 1457
1050 ..... 1066
1050 AspAlaThrPheThrAspAlaGlyThrThrAlaAspGlyTyrrThrLysI 1066
1458 C.....AATCTAGATCAGCAGCGGACGATCAAGCAAAACAAAG 1498
1066 egluThrProLeuSerAspGluInlAlaGlnAlaLeuGlyAsnGlyAspA 1083
1499 CCTTTAGTGAATCGGCTTGCTGACGGGAGGGGAGCGGTCACTGAAT 1548
1083 snSerAlaGluLeuTyrrleuThrAsp.....Asn 1092
1549 GCGGATATACGTTCAACCCCGCAACCTTATTC.....GGCTTTCG 1592
1093 AlaSerAsnAlaThrAsnGlnAspAlaSerValGlnLysProGlySerTh 1109
1593 CGGCGACGCTTGGATTGACGCGCATTCGCTTTCGATCCACCGCATTC 1642
1109 rSerPheAspLeuIleValAsnGlyGlyIlePro...AspLysIleS 1125
1643 AAAATACCGATGAAGGGCGGATGTTGTCAACCAATCAAGCAAGAA 1692
1125 erSerThrThrThrGly.....TyrrGlnAlaAsnThrGlnGly 1138
1693 TCCACCGCTTACCATACAGCAAT.....AAGATTTTCTAC 1730
1139 gLyThrTyrrThrPheSerGlyThrTyrrProAlaAlaValAspLysThr 1155
1731 AAC.....GGCAATAACAACAATCGATAGCAAAAAAGAAATG 1771
1155 rThrAspAlaGlnGlyLysLysHisAspLeuAsnThrThrTyrrAspAla 1172
1772 CTTACAAAGGTTGGTTGGCAGAAAGATCAACCAAGAAAGAAAGGCGG 1821
1172 laThrAsnSerPheThrAlaSerMetAlaValThrAsnAlaAspTyrrAla 1188
1822 CTCATCTGAATTTACCAACCGGAAGAGCGGATCGCACTTACTGCTTC 1871
1189 AlaGlnValAspLeuTyrrAlaAspLysAlaHisThrGlnLeuLeuLysH 1205
1872 CGGCGAACAATTTAAACGCAATATACGCAAAACAAAGCGCAACTGT 1921
1205 spHeAspThrLysVal..... 1210
1922 TTTTACGGCAGACGACCGCGCTTACATCTTTAGCAAGCGG 1971
1211 .....ArgLeuThrAlaProThrPheThrAspLeuLysPhe... 1222
1972 TGTCAAAATATGGAAGTATCCCAAGAGAAATGCTGTGGACACAGCA 2021
1223 .....AsnAs 1224
2022 TTGATGACCGGCACATTTAAAGCGGAAACCTTCATATTACAGCGGAC 2071
1224 nGlySerAspGlnThrSerGluAla...ThrIleLysValThrGlyThrV 1240
2072 AAGCGGTG.....GTTCCCGCAATGTTGCCAAAGTG 2103
1240 alSerSerAspThrLysThrValAsnValGlyAspThrValAlaAlaLeu 1256
2104 GAAGCGCATGCGATTAAAGCAATCACGCCCAAGCATTTTCGGTGTGCG 2153
1257 AspAlaGlnHisHisPheSerValAspValProValAsnTyrrGly... 1271
2154 ACCGCATCAAGCCACACAAATC...TGTACAGTTCGAGCTGACGCGGTC 2200
1272 .....AspAsnThrIleLysValThrAlaThrAspGluAspGlyA 1285
2201 TGACAGTTGTACGGAATAAACATTCACCGCATTAAGTATGCTTCA 2250
1285 snThrThrThrGluGlnLysThrIleThr.....Ser 1295
2251 TTGACAGACCGCATCAGACGAAATGTACGCTTCGATCCATCAGCT.. 2298
1296 SerTyrrAspProAspValLeuLysAsnAlaValThrPheAspGlnGlyA 1312
2299 .....CATTTAATCTC..... 2310
1312 lLysPheGlyAlaAsnGluPheAsnAlaThrSerAlaLysAspTyrrAsp 1329
2311 .....ACAGGACTGCTCCACACTCAACAGGC..... 2334
1339 rOLysThrGlyLeuAlaThrIleThrGlyLysValLysHisProThrThr 1345
2334 ..... 2334
1346 ThrLeuGlnValAspGlyLysGlnIleSerIleLysAsnAspLeuThrPh 1362
2335 .....AATCTTACTGACGC..... 2349
1362 eSerPheThrLeuAspLeuGlyThrLeuGlyGlnLysProPheGlyVal 1379
2350 .....GGAGACGCGCATATACGTTACGCGCAACGCAACCAAGCGC 2394
1379 alValGlyAspThr.....ThrGlnAsnLysThrPheGlnGlu 1391
2395 AACCTCAGCGCTGTGGCAATGCCCAAGCAACATTTAATCAAGCCACAT 2444
1392 AlaLeuThrPheIleLeuAspAlaValAlaPro.....ThrLeuSerIle 1406
2445 AAAGGCAACACATCGGCT.....TGGACATGCTTCATTTAATC 2485
1406 uAspSerSerThrAspAlaProValTyrrThrAsnAspProAsnPheGln 1423
2486 TAAAGCAACAGCGCGTACA.....AAC 2508
1423 lThrGlyThrAlaThrAspAsnAlaGlnTyrrLeuSerLeuAlaIleAsn 1439
2509 GGCAGTCTGACGCTTTCGACACGCTAAGGCAACGTAAGC..... 2550
1440 GlySerHisValAlaSerGlnTyrrAlaAspIleAsnSerGlyLys 1456
2551 .....CATTCGCACTGACGCAATGTCTCCCTAGCCGAT..... 2586
1456 sProGlnHisMetAlaIleAspGlnProValLysLeuGluGlyLysA 1473
2587 .....AAGCATATTCCATTTTGAACAGCGCGTTCACCGGA 2625
1473 snValLeuThrValAlaValThrAspSerGluAsnAsnThrThrLys 1489
```

2626 A A A A T C A G C G G C G A G A T A C G C A T T A C A C T T A A A A G A C A G G A A T G 2675  
 1490 L y s I l e T h r V a l T y r T y r G l u P r o l y s t h r L e u A l a I a l P r o T h V a 1506  
 2676 G A C G C T C C G T C G G C A G C G A A T T A G C A A T T T A A C C T T G C A C A G C C A 2725  
 1506 I l h r . . . P r o S e r T h r T h r G l u P r o A l a l y s . . . . . T h r V a l T 1518  
 2726 C C A T T A C A C T C A A T T C G C G T T A C A C A G A T G C G G C A G G C G C A A A C C 2775  
 1518 h l e u T h r I n l y s A l a s n A l a l a . . . . . A l a t h r G l y l u T h r V a l 1530  
 2776 G C A C T G C G C G A G A T G C G C G C G C G C G C G C G C G C G C G T T A T T A T C 2825  
 1531 G l n T y r S e r A l a s p l y G l y l y s t h r T y r G l n s p a l P r o A l a l a g l 1547  
 2826 C G T T A C G C G C C A C T T G G C A G A A T C C G T T T A C A C G C T G A C G T A A 2875  
 1547 y a l . . . . . T h r V a l T h r A l a 1553  
 2876 A C G C A A T T G A A C G T C A G G A C A T T C G C T T A T G T G . . . G A A C T C 2922  
 1553 s n . . . . . G l y t h r P h e l y s P h e l y S e r T h r A s p l e u 1563  
 2923 T T C G G C T A C C G C A G C G A A A T T G A A C T G C G G A A A C T T C G A A G C A C 2972  
 1564 T y r G l y . . . . . A s n G l u S e r P r o A l a V a 1571  
 2973 T T A C A C C T T G G C T G C A C A T A C C G C A C A A C C C G T A G T C G A C 3022  
 1571 I a s p t y r V a l T h r s n I l e l y s A l a s p a s p r o A l a g l n l e u g l n . 1587  
 3023 A A T T A C G C T A G T G A A G A A A A C A C A C A C C G T G T C C G A A A A T C T T 3072  
 1588 . . . . . T h r A l a l y s e l n A l a l e u t h r a s n l e u I l e a l a s e r . . . . . 1599  
 3073 A A T T T C A C C C T G C A A A C A A C A C G T G A T G C C G C G C A T G C C T T A T C A 3122  
 1599 . . . . . 1599  
 3123 G C T T A T C C G A A A G A C G G C A G T C C G C T G C A T A T C C G T C A A A G A C 3172  
 1600 . . . . . A l a l y s t h r l 1603  
 3173 A A G A G C T T C C G A A A C T C G C A A G C G G A G A A A C A G A G C C G C C T T G 3222  
 1603 e u s e r A l a s e r G l y l y s t y r A s p a s p a l a t h r T h r A l a l e u A l a l a 1619  
 3223 A C G C A A A A C G A C A A C T T G C C G C C A A C A C A G C G G A A A A A C A C A A 3272  
 1620 A l a t h r G l n l y s A l a g l n T h r A l a l e u a s p l n . . . . . T h 1631  
 3273 C G C G A A A G C C T T G A C G C G T G A T T G C G C G G C G C A T G C C A C C G A A A 3322  
 1631 r a s p a l a s e r V a l a s p e r l e u t h r G l y A l a s n A r g A s p l e u g l n T h r A 1648  
 3323 A G C A G A A A A G T G T T C C G A A C C G C C G C A G C A G C G G G A A A A T G C C 3372  
 1648 I a l l e a s n g l n l e u A l a . . . . . 1653  
 3373 G G C A T T A T G A G G C G A G A A A A A A A A A A C G G T C A G C G C G A T A A A G A 3422  
 1654 . . . . . A l a l y s l e u P r o A l a s p l y s t y 1661  
 3423 C A C C G C C T T G C G A A A C A G C G C G A A A C C C G C G C T A C C A C C G 3472  
 1661 s t h r S e r l e u a s n g l n l e u s l n s e r V a l y s A l a l a l e u g l y T h r . 1677  
 3473 C C T T C C C C C C C C C C C C C C C C G C G G A T T G C C G A C C G A C G A C C C 3522  
 1678 . . . . . A s p l e u . G l y A s n g l n T h r . . . . . 1683

3523 C A C C G A C A C C C A C A C C G C A C C G A C C T G A T A C C C T T A T G C A A T A G 3572  
 1684 . . . . . A s p p r o s e r t h r G l y l y s t h r . . . . . 1690  
 3573 C G G T T T A G T A A T T T T C C C C A C G C T C A A C A C A G C G T T T G C C G T A C A G G 3622  
 1690 . . . . . 1690  
 3623 A C G A A T T G A C C G C G T T T G C C G A A G A C C C G C A A C C G C G T T T G A C A 3672  
 1691 . . . . . P h e r t h r A l a l e u A s p a s p l e u V a l a l a . . . . . G l 1701  
 3673 A G G G C A T C C G G A C A C A C A C A C T A C C G T T G C A A G A T T T C G C C C T A 3722  
 1701 n A l a g l n A l a g l y T h r G l n t h r A l a s p l n l e u g l n A l a s e r l e u A l a l 1718  
 3723 C C G C A A A A A C C G A C C T G C C C A A T C G T A T G C A A A A A C T T G C G A 3772  
 1718 y s v a l l e u s p a l a l a l e u A l a l y s . . . . . l e u A l a 1728  
 3773 G C G G C G C G C G C A T C C T G T T T G C A C A C C G A C G C G A A A C A C C T T C 3822  
 1729 G l u G l y I l e l y s A l a l a . . . . . T h r P r o A l a g l u V a l a l y s 1741  
 3823 G A C G A G C G A T G C G A C T G C G A C G C T T G C C A C G T G C C G T T T C G 3872  
 1741 n A l a l y s A s p A l a l a t h r . . . . . G l y l y s t h r T y r I r a l a s p l l e a 1756  
 3873 G C A A T A C G C A T C G C A G G T T G C A T C G C A T C A C A C C G C G C G G G T T 3922  
 1756 I a s p t h r l e u t h r s e r G l y n A l a s e r A l a s p a l a s e r A s p l y s l e u 1772  
 3923 T T T A G T A C C G C C C T T T C A G A C G C A T C A G A G C A A A A T C C G C C G C C 3972  
 1773 . . . . . A l a h i s l e u g l n A l a l e u g l n S e r l e u T h r l y s V a l A l 1786  
 3973 G T G C T G A T T A C G C A T T C A G C A A . . . . . 3997  
 1786 a l a l a l e u A l a g l u A l a l a l y s t h r A l a g l y l y s e l y s a s p t h r t h r g 1803  
 3997 . . . . . 3997  
 1803 I y t h r s e r A s p l y s e l y g l y g l n g l y T h r P r o A l a P r o A l a P r o g l y 1819  
 3998 G A T A C C G C G C A G T T T C G C G G A T T C G G A T C G A C C G A C C A C T G G C C A 4047  
 1820 A s p t h r G l y l y s A s p l y s e l y s p e l n g l y s e r l n P r o s e r S e r G l y G l 1836  
 4048 A C C G C T A T T T G T C C A A A A C G G A T T A C G A T A C G A A A A C G T C A A T A T 4097  
 1836 y a s n l l e P r o t h r l y s P r o A l a t h r T h r S e r t h r s e r t h r A s p a s p t 1853  
 4098 C G C A C C C C G G C C T T G C A T T A C C G C T A C C G C G C G C G A T T A A G C A G 4147  
 1853 h t t h r A s p t r g a s n g l y g l n l s t h r S e r G l y l y s G l y A l a l e u P r o l y s 1869  
 4148 A T T A T T C A T T C A A C C G C G C A C A C A T T T 4177  
 1870 T h r A l a g l u T h r T h r G l u a r g P r o A l a p h e 1879

seq\_name: SwissProt\_40:STFR\_ECOLI

seq\_documentation\_block:  
 ID STFR\_ECOLI STANDARD; PRT; 1120 AA.  
 AC P76072; P77560;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Side tail fiber protein homolog from lambdaoid prophage Rac.  
 GN STFR OR B1372.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.



```
4037 .....ACATC 4041
422 AlaGlySerAlaThrAlaAlaIaGlnSerLysSerThrAlaGlnSerAl 438
4042 GGGGCAACGGCGTATTTCGCCAAAAGCGATACCGATACGAAACGT 4091
438 aaIaThrArgAlaGlnThrAlaAlaIaLysArgAlaGlnAspIleAlaSerA 455
4092 CAATATCGCCACCCGGCCCTGCATTCACCGCTACCGCGGCGCATTA 4141
455 IVal.....AlaLeuGluAspAlaSerThrThrLys..... 465
4142 AGCGCATTTATTCATTCAAACCGCGCAACACATTCATCAGCCTTAT 4191
466 LysGlyIleValGlnLeuSerSerAlaThrAsnSerThrSerGluThrIle 482
4192 TTGAGCGCTGCTCTATACGATGCGGCTTCGGCAAGTCCGAAAGCGCGT 4241
482 uAlaIaIaThrPro.....LysAlaValLysSerA 492
4242 CAATACCGCGGTATTGGCGCAGGATTCGGCAAAACCGCAGTGGCAAT 4291
492 IaTyAspAsnAlaGlnLysArgLeuGlnLysAsp.....GlnAsn 505
4292 GGGGCGTAACCCGCAA.....TCAAGGT 4317
506 GlyAlaAspIleProAspLysGlyCysPheLeuAsnIleAsnAlaVal 522
4318 TTCACGCTGCTCCCTCCACGCTGCCCGCGCAAGGGCGCGCAATTGGAAGC 4367
522 IserLys.....ThrAspPheAlaAspLysArgGlyMetArgIleValA 537
4368 GCAGCACAGCGGGCATCAATTAGCTACCGCTGCT 4405
537 rgValAsnAlaProAlaGly.....AlaThrSerGly 547
```